Pro Gradu

An \textit{ab initio} Investigation of Sulfuric Acid Monohydrate Formation

L. J. Partanen

8.8.2011

University of Helsinki
Department of Chemistry
Laboratory of Physical Chemistry
An ab initio Investigation of Sulfuric Acid Monohydrate Formation

Sulfuric acid and its gaseous hydrates play a central role in the nucleation processes of the atmosphere. In this study, the thermodynamic properties for the formation of the sulfuric acid monohydrate complex were determined from the results of accurate ab initio calculations by using statistical mechanics. Of the ab initio calculations, geometry optimizations and energy calculations were performed with the explicitly correlated CCSD(T)-F12a/VDZ-F12 method, which was shown to give results comparable to CCSD(T)/aug-cc-pVQZ level calculations. Four different stable geometries were found, and the energies of the two lowest were within 0.41 kJ mol⁻¹ of each other. Harmonic frequencies were calculated both at the DF-SCS-LMP2/aug-cc-pVTZ level and the CCSD-F12/VDZ-F12 level.

Because the harmonic approximation works badly for the high frequency OH stretches and the low frequency intermolecular large amplitude motions, some of the vibrational degrees of freedom were treated by limiting the dimensionality of the potential energy surface to small, two- or three-dimensional subspaces that contained a few strongly coupled vibrational degrees of freedom. In these anharmonic domains, the vibrational problem was solved variationally from a potential energy surface calculated at the CCSD(T)-F12a/VDZ-F12 level. Even though the subspaces are completely decoupled from the rest of the vibrational degrees of freedom, my results indicate that with a careful choice of the domains, the resulting vibrational states are accurate enough for the calculation of thermodynamic properties.

It is shown that with the anharmonic domain approximation it becomes possible to incorporate quantum mechanically the presence of at least some of the other minimum energy structures into the thermodynamic properties, and that the inclusion of these is essential if accurate results are desired. Furthermore, the anharmonic domain approximation makes it relatively easy to calculate vibrational overtones which, especially for the large amplitude motions, have a major impact on the vibrational partition function.

With the inclusion of the anharmonic domains, very uniform results were obtained for the thermodynamic properties with the two different methods used in the harmonic calculation. At one atmosphere and 298 K, the Gibbs free energy was found to have a value of about -8.0 kJ mol⁻¹. The anharmonic domains used in this study had the effect of raising the zero point energy by around 1.7 kJ mol⁻¹. Comparison with the earlier results reveals that one of the most important reasons for the differences in the Gibbs energy is the inaccurate calculation of the electronic energies of the different reaction components. Thus, we recommend that all future studies employ a higher level method for these calculations.

Finally, investigations were carried out on the temperature dependence of the equilibrium constant and on the pressure and temperature dependencies of the Gibbs free energy and entropy. To this effect, the statistical thermodynamic formulae that included the anharmonic domain approximation in calculations of the thermodynamic properties were derived. The temperature dependence of both enthalpies and entropies was predicted to be rather small and, therefore, the Gibbs free energy varied almost linearly with temperature.

Avainsanat – Nyckelord – Keywords

sulfuric acid monohydrate, anharmonic domain, thermodynamic property, sulfuric acid

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aug</td>
<td>augmented basis set</td>
</tr>
<tr>
<td>aug-cc-pV mZ</td>
<td>augmented correlation consistent polarized valence Double ($m=D$), Triple ($m=T$), Quadruple ($m=Q$) or pentuple ($m=5$) Zeta basis set with diffuse functions</td>
</tr>
<tr>
<td>aug-cc-pV mZ*</td>
<td>augmented correlation consistent polarized valence Triple ($m=T$) or Quadruple ($m=Q$) Zeta basis set with correlation consistent polarized valence Triple ($m=T$) or Quadruple ($m=Q$) Zeta basis set used for the hydrogens</td>
</tr>
<tr>
<td>aug-cc-pV(T+d)Z</td>
<td>augmented correlation consistent polarized valence Triple Zeta basis set with an extra tight d function for the third row atoms</td>
</tr>
<tr>
<td>AV mZ</td>
<td>augmented correlation consistent polarized valence Double ($m=D$), Triple ($m=T$), Quadruple ($m=Q$) or Pentuple ($m=5$) Zeta basis set with diffuse functions</td>
</tr>
<tr>
<td>AVDZ/MP2FIT</td>
<td>auxiliary augmented correlation consistent polarized valence Double Zeta basis set</td>
</tr>
<tr>
<td>BLYP</td>
<td>a density functional</td>
</tr>
<tr>
<td>B3LYP</td>
<td>a hybrid density functional</td>
</tr>
<tr>
<td>BSSE</td>
<td>basis set superposition error</td>
</tr>
<tr>
<td>cc</td>
<td>correlation consistent</td>
</tr>
<tr>
<td>cc-pV(T+d)Z</td>
<td>correlation consistent polarized valence Triple Zeta basis set with an extra tight d function for the third row atoms</td>
</tr>
<tr>
<td>CC</td>
<td>Coupled Cluster</td>
</tr>
<tr>
<td>CCS</td>
<td>Coupled Cluster singles</td>
</tr>
<tr>
<td>CCSD</td>
<td>Coupled Cluster singles and doubles</td>
</tr>
<tr>
<td>CCSD-F12a</td>
<td>explicitly correlated Coupled Cluster singles and doubles with some of the contributions from the explicitly correlated configurations to the doubles residual neglected.</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>Coupled Cluster singles and doubles with perturbative triples</td>
</tr>
<tr>
<td>CCSD(T)-F12</td>
<td>explicitly correlated Coupled Cluster singles and doubles with perturbative triples</td>
</tr>
<tr>
<td>CCSD(T)-F12a</td>
<td>explicitly correlated Coupled Cluster singles and doubles with perturbative triples and with some of the contributions from the explicitly correlated configurations to the doubles residual neglected.</td>
</tr>
<tr>
<td>CC-VSCF</td>
<td>correlation corrected vibrational self-consistent field</td>
</tr>
<tr>
<td>CGTO</td>
<td>contracted Gaussian type orbitals</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Name/Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>CP</td>
<td>counterpoise</td>
</tr>
<tr>
<td>CSF</td>
<td>configuration state function</td>
</tr>
<tr>
<td>DF-SCS-LMP2</td>
<td>density fitted spin component scaled local second order Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>DZ</td>
<td>Double Zeta</td>
</tr>
<tr>
<td>DZP</td>
<td>polarized Double Zeta basis set</td>
</tr>
<tr>
<td>GTO</td>
<td>Gaussian type orbitals</td>
</tr>
<tr>
<td>G2</td>
<td>Gaussian – 2 composite extrapolation technique</td>
</tr>
<tr>
<td>G3</td>
<td>Gaussian – 3 composite extrapolation technique</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree–Fock</td>
</tr>
<tr>
<td>LMP2</td>
<td>local second order Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>MP</td>
<td>Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>MP_m</td>
<td>mth order Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>PES</td>
<td>potential energy surface</td>
</tr>
<tr>
<td>PW91</td>
<td>a density functional</td>
</tr>
<tr>
<td>QZ</td>
<td>Quintuple Zeta</td>
</tr>
<tr>
<td>RMS</td>
<td>root mean square</td>
</tr>
<tr>
<td>STO</td>
<td>Slater type orbital</td>
</tr>
<tr>
<td>VDZ</td>
<td>valence Double Zeta basis set</td>
</tr>
<tr>
<td>VDZ-F12</td>
<td>valence Double Zeta basis set optimized for explicitly correlated methods</td>
</tr>
<tr>
<td>VDZ/JKFIT</td>
<td>auxiliary valence Double Zeta basis set</td>
</tr>
<tr>
<td>VDZ-F12/optri</td>
<td>auxiliary valence Double Zeta basis set for the resolution of identity</td>
</tr>
<tr>
<td>V(T+d)Z</td>
<td>valence Triple Zeta basis set with extra tight d function for the third row atoms</td>
</tr>
<tr>
<td>ZPE</td>
<td>zero point energy</td>
</tr>
<tr>
<td>6-311++G(2d,2p)</td>
<td>doubly polarized valence triple zeta basis set with diffuse functions on all atoms</td>
</tr>
<tr>
<td>6-311++G(3df,3pd)</td>
<td>triply polarized valence triple zeta basis set with diffuse functions on all atoms</td>
</tr>
</tbody>
</table>
## Contents

1 Introduction 1

2 The Theoretical Basis 3
   2.1 Sulfuric Acid in the Atmosphere 3
      2.1.1 Sources of Sulfuric Acid 3
      2.1.2 Sulfuric Acid and Cloud Formation 5
   2.2 Quantum Chemical Principles 6
      2.2.1 The Born-Oppenheimer and Adiabatic Approximations 7
      2.2.2 The Hartree–Fock Approach 9
      2.2.3 Electron Correlation 11
      2.2.4 MP2 12
      2.2.5 Coupled Cluster Methods 19
      2.2.6 Basis Functions 25
      2.2.7 The Solutions to the Nuclear Hamiltonian 27
   2.3 Rudimentary Statistical Thermodynamics 32
      2.3.1 Ensembles and Partition Functions 32
      2.3.2 The Evaluation of Partition Functions 35
      2.3.3 Statistical Mechanics of Equilibrium Systems 37

3 The Computational Results 40
   3.1 Geometry Optimizations and Energies 40
      3.1.1 Water and Sulfuric Acid Molecules 40
      3.1.2 Sulfuric Acid Monohydrate 42
   3.2 Harmonic Vibrational Wavenumbers 45
      3.2.1 Water and Sulfuric Acid Molecules 45
      3.2.2 Sulfuric Acid Monohydrate 47
   3.3 Anharmonic Vibrational Wavenumbers 49
      3.3.1 The Three-Dimensional H$_2$O Problem 49
      3.3.2 The Two-Dimensional Large Amplitude Problem 52
   3.4 Partition Functions 56
   3.5 Thermodynamic Parameters 58
      3.5.1 Temperature and Pressure Dependences of the Parameters 63
   3.6 Summary and Discussion 66
1 Introduction

Sulfuric acid is one of the dominating products in chemical industry. Equally importantly, it plays a central part in environmental chemistry due to its role as one of the key components in acid rain and in cloud formation.\(^1,^2\) Cloud formation occurs via condensation of water into so-called condensation nuclei, which are thought to be complexes of sulfuric acid with water, ammonia and others.\(^3^--^5\) The fact that most of the sulfuric acid in the atmosphere is contained in hydrates has a huge impact on the nucleation models that seek to predict the nucleation rates.\(^6\) To estimate accurately the effect of the hydrates, knowledge of the equilibrium constants for the hydration reactions is necessary.\(^7\)

Experimentally the equilibrium constants for the individual hydrate reactions are difficult to find because it is hard to estimate the amounts of the different hydrates formed upon reaction of water with sulfuric acid.\(^8\) Thus, the standard approach is to use quantum chemical calculations to obtain the equilibrium constants. Due to the large size of many of the complexes, most of the computational studies have until now used second order Møller–Plesset Perturbation theory (MP2) or density functional (DFT) methods.\(^9^--^13\)

Recently, major developments have been achieved in the different explicitly correlated methods that take into account the singularities in potential energy at points where two electrons coincide.\(^14^--^18\) Specifically, the explicitly correlated Coupled Cluster method with perturbative triples and some additional approximations to the explicitly correlated treatment, CCSD(T)-F12a, has made it possible to obtain very accurate electronic energies and equilibrium structures with substantially less computational time than required in the standard CCSD(T) methods.\(^18,^19\)

In MP2 calculations, the use of localized orbitals has made it possible to ignore two-electron integrals between distant electrons, and to restrict the number of virtual orbitals any given electron can be excited to.\(^20\) These local approximations radically reduce the computational scaling with respect to system size and are virtually free of the basis set superposition error (BSSE).\(^20^--^22\) Additional reductions in computational time of about an order of magnitude can be achieved by introducing density-fitting approximations to the many two electron integrals in MP2. Furthermore, to account for the biased description of singlet and triplet electron pairs, the spin component scaling (SCS) methods separately scale the singlet and triplet MP2 energies, so that a more uniform distribution of results is achieved.\(^24^--^26\)

The statistical mechanical formulae used for the calculation of thermodynamical properties requires the knowledge of electronic energies, vibrational states and rotational constants.\(^27,^28\) Because the full anharmonic calculation for the sulfuric acid monohydrate and sulfuric acid molecules is an overwhelming task due to the computational demands of the variational calculation, harmonic approximation is thought necessary to obtain the vibrational states. Previously, the effects of using different computational methods for the harmonic frequencies have been studied\(^29\) and even the effect of anharmonic frequencies has been considered with methods such as the correlation corrected vibrational self-consistent field (CC-VSCF).\(^11,^30\)
The purpose of this study is to accurately calculate the equilibrium constant and other thermodynamic properties for the reaction of sulfuric acid and water to form a sulfuric acid monohydrate complex. I have opted to use the CCSD(T)-F12a method with the valence double zeta VDZ-F12 basis set for the calculation of energies and equilibrium structures. For the harmonic vibrational frequencies I used either the same VDZ-F12 basis combined with a simplified CCSD-F12a method, where the perturbative triplets are missing, or a density fitted local spin component scaled MP2 (DF-SCS-LMP2) calculation with the Dunning augmented correlation-consistent polarized valence triple zeta basis set (aug-cc-pVTZ).

To ascertain the validity of my results, I have widely compared them with those of past computational studies and experimental results for equilibrium geometries, harmonic frequencies and, of course, the thermodynamic properties themselves. In large, relatively weakly bound complexes such as the sulfuric acid monohydrate, the harmonic approximation often fails to predict the large amplitude motions corresponding to intermolecular motions of the component molecules. On the other hand, it is also known to fail for the high-frequency OH stretching vibrations. Because of this, all research striving for high accuracy in such systems must find some way to deal with anharmonicity. In this study, anharmonic effects were considered by limiting the dimensionality of the potential energy surface (PES) to small, two- or three-dimensional subspaces that contained a few strongly coupled vibrational degrees of freedom. For these anharmonic domains, the vibrational problem was solved variationally. Because the domains are completely uncoupled to the rest of the vibrational degrees of freedom, great care was taken to confirm the reliability of results with experimental data and other anharmonic calculations when experimental data was not available. As cloud formation occurs mainly in troposphere at a temperature range from around 210 to 300 K, the temperature dependence of the equilibrium constant was studied. To obtain a general picture of the hydrate formation process as a function of altitude, the pressure and temperature dependencies of the reaction Gibbs free energy and reaction entropy were investigated.
2 The Theoretical Basis

2.1 Sulfuric Acid in the Atmosphere

2.1.1 Sources of Sulfuric Acid

The origins of sulfur acid in the atmosphere are manifold. A number of compounds containing sulfur in the lowest oxidation state such as hydrogen sulfide, carbonyl sulfide, carbon disulfide, methyl mercaptan, dimethyl sulfide, and dimethyl disulfide can function as precursors for sulfuric acid. These compounds are released to the air from the seas and as byproducts from the reactions occurring within microbiological organisms.\(^2\)

In many cases, the first step in the atmospheric transformation of these substances into \(\text{H}_2\text{SO}_4\) is the oxidation of sulfur and the formation of \(\text{SO}_2\). The oxidation processes are usually complex, but oxidation via a reaction with the hydroxyl radical is often an important contributor to the oxidation process as a whole.\(^1\) For the simplest case of \(\text{H}_2\text{S}\), the reaction is

\[
\text{H}_2\text{S} + \cdot \text{OH} \rightarrow \text{H}_2\text{O} + \cdot \text{SH},
\]

and it results in the production of a thionyl radical. This radical quickly reacts further to give sulfur dioxide: \(^{45-47}\)

\[
\cdot \text{SH} + \text{O}_2 \rightarrow \text{SO} + \cdot \text{OH}
\]

\[
\cdot \text{SH} + \text{O}_3 \rightarrow \text{SHO} \cdot + \text{O}_2
\]

\[
\text{SHO} \cdot + \text{O}_2 \rightarrow \text{SO} + \text{HOO}^-\]

\[
\text{SO} + \begin{cases} 
\text{O}_2 \\
\text{O}_3 \\
\text{NO}_2
\end{cases} \rightarrow \text{SO}_2 + \text{P},
\]

where \(\text{P}\) denotes simply the other products formed.\(^2\)

Most of the sulfuric acid in the atmosphere arises from sulfur dioxide. In addition to the \(\text{SO}_2\) formed from the atmospheric sulfur-containing compounds, around 90% of the sulfur in fossil fuels is released to the atmosphere directly as \(\text{SO}_2\).\(^1\) Because fuel combustion and industrial activities make up around 76% of the global emissions of sulfur compounds, these direct emissions turn out to be the dominating source for \(\text{SO}_2\). Significant amounts of \(\text{SO}_2\) are also released from volcanoes and underwater fissures.

The oxidation of \(\text{SO}_2\) to sulfuric acid can in principle occur homogenously within the gas phase, heterogenously within liquid droplets or on the surface of aerosols. The rate of the process and its mechanisms are highly dependent on the nature and presence of an aqueous phase, the concentration of oxidizing species like \(\text{H}_2\text{O}_2\) and \(\text{O}_3\) and the intensity of solar radiation. Below is given an elaboration of each of the pathways.

In the gas phase, the only important way the oxidation of sulfur dioxide occurs is through a reaction with a hydroxyl radical. The reaction has a rate determining step\(^1\)

\[
\text{SO}_2 + \cdot \text{OH} + \text{M} \rightarrow \text{HOSO}_2 + \text{M}.
\]
From the H\textsubscript{2}SO\textsubscript{2} species, there exists a series of rapid reactions that can lead to sulfuric acid. The most important of these is the reaction with oxygen yielding SO\textsubscript{3}:

\[
\text{H\textsubscript{2}SO\textsubscript{2}} + \text{O}_2 + \text{M} \rightarrow \text{HOO} \cdot + \text{SO}_3 + \text{M}.
\] (7)

When sulfur trioxide dissolves into water the product is sulfuric acid

\[
\text{SO}_3 + \text{H}_2\text{O} \rightarrow \text{H}_2\text{SO}_4.
\] (8)

Reaction (8) is thought to occur via the formation of an SO\textsubscript{3} \cdot H\textsubscript{2}O complex which then reacts with another water molecule to give sulfuric acid as the product.\textsuperscript{49,50} The hydroxyperoxyl radical formed by reaction (7) may react further with nitrogen monoxide in a reaction which regenerates the hydroxyl radical

\[
\text{HOO} \cdot + \text{NO} \rightarrow \cdot \text{OH} + \text{NO}_2,
\] (9)

and gives rise to a possible chain reaction.\textsuperscript{2} In addition the generation of HOO\cdot leads to an increased formation of the oxidizing agent H\textsubscript{2}O\textsubscript{2}.\textsuperscript{6} Of the other radicals that could contribute to the gas-phase oxidation, the Criegee biradical is probably most important.\textsuperscript{51,52} However, when taking into account the heterogenous oxidation pathways, it is unlikely to have much of an effect on the total rate of oxidation.\textsuperscript{1}

The heterogenous pathway from SO\textsubscript{2} to sulfuric acid begins by the solvation of sulfur dioxide and its reaction in the water solution

\[
\text{SO}_2(\text{g}) \rightleftharpoons \text{SO}_2(\text{aq})
\] (10)

\[
\text{SO}_2(\text{aq}) + 2 \text{H}_2\text{O} \rightleftharpoons \text{HSO}_3^- (\text{aq}) + \text{H}_3\text{O}^+ (\text{aq})
\] (11)

\[
\text{HSO}_3^- (\text{aq}) + \text{H}_2\text{O} \rightleftharpoons \text{SO}_3^{2-} (\text{aq}) + \text{H}_3\text{O}^+ (\text{aq}).
\] (12)

The equilibria in reactions (10) through (12) is quickly attained,\textsuperscript{1} and it follows that in water SO\textsubscript{2} exists in three different forms, all of which show different behaviour with different oxidizing agents. According to Le Châteliers principle, the concentration of SO\textsubscript{3}\textsuperscript{2-} (aq) increases as pH increases. It also follows from this that the higher the pH the greater the total amount of different sulfur species in solution. Possible oxidizing agents in the liquid phase include nitric oxides and free radicals, but the most important ones are O\textsubscript{2}, O\textsubscript{3} and H\textsubscript{2}O\textsubscript{2}.\textsuperscript{2}

While the oxidation by O\textsubscript{2} in pure water solution is too slow to have a significant impact on the total oxidation of sulfur dioxide, the Fe\textsuperscript{3+} and Mn\textsuperscript{2+} ions, present in small amounts within the droplet, catalyze the reaction. Possible mechanisms have been studied, but they are exceedingly complex and will not be presented here.\textsuperscript{53-56} All of the suggested mechanisms predict that the overall rate constant increases as a function of pH. This means that they form a self-deactivating cycle, as the acidic product of these reactions has the effect of lowering the pH of the solution.

A similar self-deactivation where the rate constant increases with the pH of the solution is observed when ozone acts as an oxidizing agent. Because it is abundant in the atmosphere, there usually exists sufficient amounts of ozone dissolved in liquid droplets to have a significant impact on the oxidation of SO\textsubscript{2}, in spite of the low value of its Henry’s law constant of 1.1 \cdot 10\textsuperscript{-2} mol dm\textsuperscript{-3} atm\textsuperscript{-1}.\textsuperscript{57} Ozone reacts in a unique way with each of the different sulfur dioxide species as represented in the following reactions
SO$_2$(aq) + O$_3$(aq) + H$_2$O $\rightarrow$ H$_2$SO$_4$(aq) + O$_2$ \hfill (13)

HSO$_3^-$ (aq) + O$_3$(aq) $\rightarrow$ HSO$_4^-$ (aq) + O$_2$ \hfill (14)

SO$_3^{2-}$(aq) + O$_3$(aq) $\rightarrow$ SO$_4^{2-}$(aq) + H$_2$O + O$_2$. \hfill (15)

The precise mechanisms underlying these reactions were originally proposed by Hoffman.\textsuperscript{58}

Because of its high Henry’s law constant of 1·10$^5$ mol dm$^{-3}$ atm$^{-1}$,\textsuperscript{57} hydrogen peroxide can be present in water droplets in relatively large quantities even if its gas-phase concentration is small. The reaction between H$_2$O$_2$ and SO$_2$ occurs in two steps.\textsuperscript{2} The first step is

\[
\text{HSO}_3^-(aq) + \text{H}_2\text{O}_2(aq) \rightleftharpoons \text{HOOSO}_2^-(aq) + \text{H}_2\text{O}
\] \hfill (16)

followed by a reorganization and protonation of the ion

\[
\text{HOOSO}_2^-(aq) + \text{H}_3\text{O}^+ \rightarrow \text{H}_2\text{SO}_4(aq) + \text{H}_2\text{O}.
\] \hfill (17)

It is thought that organic peroxides could oxidize SO$_2$ in a similar manner. However, due to their low mixing ratios and Henry’s law constants when compared to H$_2$O$_2$, these reactions are found to contribute only a little to the oxidation when compared to hydrogen peroxide.\textsuperscript{1}

The key difference between the oxidation by hydrogen peroxide and O$_2$ or O$_3$ is that the rate constant of the hydrogen peroxide depends inversely on pH. This means that a decrease in proton concentration results in a decrease in the rate constant, while increasing the available sulfur species in the usual manner. As a result, the rate of the peroxide oxidation stays relatively constant in pH 1 − 5. It is this fact combined with the high Henry’s law constant and high reactivity of H$_2$O$_2$ that make the peroxide pathway the most important of all the oxidation pathways in this pH region. As the pH increases the ozone and different oxygen pathways become predominant.\textsuperscript{6}

The importance of oxidation on aerosol surfaces is difficult to assess, because there are so many factors at play. The relative rates depend on the physico-chemical nature of the surfaces including surface defects, surface areas and the presence of other adsorbed species. So far, studies have been able to show that oxidation on surfaces does take place and may have a significant impact on the total oxidation of the various sulfur dioxide species.\textsuperscript{1,59}

### 2.1.2 Sulfuric Acid and Cloud Formation

Once sulfuric acid is formed, there exists a number of different mechanisms through which it can react further. Because the possible reactions vary largely from photodissociation to reactions with a number of airborne species to form different kinds of compounds, sulfuric acid gives rise to a wealth of different environmental effects. Like all aerosols, the particles formed by the reaction of the sulfuric acid scatter light and have a profound effect on the world’s climate. They also indirectly affect climate via their function as condensation nuclei in cloud formation because clouds both reflect and absorb the radiation from the sun, and the thermal radiation of the earth.

In a gas containing water vapor, sulfuric acid reacts with water via

\[
\text{H}_2\text{SO}_4(g) + \text{H}_2\text{O}(g) \rightarrow \text{H}_2\text{SO}_4 \cdot \text{H}_2\text{O}(g).
\] \hfill (18)
Equation (18) is an example of binary homogenous nucleation reaction, in which two compounds merge in the gas phase. Usually this reaction is followed by further reactions with water

\[ H_2SO_4 \cdot (H_2O)_{n-1}(g) + H_2O(g) \rightarrow H_2SO_4 \cdot (H_2O)_n(g) \]  

(19)

and with additional sulfuric acid molecules, ammonia and others, forming a nucleation process. The end result of this nucleation is the birth of a premucleation embryo, upon which water can start accumulating through condensation. Eventually this can lead into the formation of raindrops and entire clouds.

It was shown by Doyle\(^6\) that extremely small amounts of sulfuric acid is capable of inducing nucleation, even if the relative humidity is less than 100 \%. As a result, the reaction system of (18) and (19) has turned out to be the most important binary nucleation process in the atmosphere\(^6\) and has received a lot of scientific attention. Investigations have also been carried out on the other possible components of the nucleation process including ions,\(^3\) ammonia\(^4\) and organic compounds.\(^5\)

One of the key goals in atmospheric research is the accurate determination of the nucleation rates, i.e., the rates at which the condensation nuclei are formed. In terms of classical nucleation theories,\(^6\) the sulfuric-acid water system is problematic as the existence of the already hydrated H\(_2\)SO\(_4\) species has to be taken into account.\(^6\) This is because the formation energy of the hydrates is negative, and so it is much more difficult to form clusters out of them than from pure monomers.\(^12\) As a result, it is important to know how much of the sulfuric acid is free relative to the total amount available. This can be obtained from equation\(^7\)

\[ \frac{\rho_{a}^{\text{total}}}{\rho_{a}^{\text{free}}} = 1 + \sum_{i=1}^{N} \prod_{j=1}^{i} K_j \left( \frac{\rho_{w}^{\text{free}}}{\rho_0} \right)^i, \]  

(20)

where \(\rho_{a}^{\text{total}}\) is the total concentration of sulfuric acid in the gas phase, \(\rho_{a}^{\text{free}}\) is the concentration of free sulfuric acid molecules, \(\rho_{w}^{\text{free}}\) is the concentration of free water molecules, and \(N\) is the number of water molecules for the largest hydrate considered. The equilibrium constants \(K_j\) correspond to the reactions (18) and (19). The \(\rho_0\) term in the denominator is the reference vapor concentration \(\rho_0 = p_0/kT\), where \(p_0\) is usually chosen as 1 atm.

Because most of the sulfuric acid in the atmosphere is hydrated, inclusion of hydrate formation into the nucleation rate models can reduce the rates by a factor of \(10^5 - 10^6\).\(^6\) In all nucleation models for the sulfuric acid system, the addition of hydrates requires the knowledge of the \(K_j\) equilibrium constants, and so it is imperative that these are known with great accuracy. Due to their complexity, the exact formulae for the sulfuric-water nucleation will not be represented here, but can be found in the literature.\(^7,62,63\)

### 2.2 Quantum Chemical Principles

In quantum chemistry, the state of the system is completely described by the wavefunction \(\Psi\). This means that from the wavefunction the expectation values of energies, particle locations and all other physical properties can be deduced. It follows that quantum chemical calculations generally revolve
around finding the wavefunction and then obtaining the desired properties from that. The system’s wave function is found by solving its Schrödinger equation

$$\hat{H}\Psi = i\hbar \frac{\partial \Psi}{\partial t}, \quad (21)$$

where $\hat{H}$ is the system’s Hamiltonian, i.e. its energy operator, $\hbar$ is the Planck’s constant $\hbar$ divided by $2\pi$ and $i$ is the imaginary unit. In this equation, the wave function depends on the locations of all particles in the system and time $\Psi = \Psi (x_1, x_2, \ldots, x_n, t)$.

Often we are interested in systems where the probabilistic aspects of the wavefunction do not vary with time. In these stationary states, equation (21) can be separated into different parts that represent the time and space variations of $\Psi$. Consequently, it becomes possible to partition the wavefunction into its time and space components $\Psi = \psi (x_1, x_2, \ldots, x_n) \tau (t)$, where the time dependence is simply $\tau (t) = \exp(-iEt/\hbar)$ and $E$ is the system’s energy. With this notation, the time-independent Schrödinger equation has the following form

$$\hat{H}\psi = E\psi. \quad (22)$$

In essence, it is the eigenvalue equation of the system’s energy. If we ignore relativistic effects,\(^6\) the Hamiltonian operator in equation (22) consists of the operators for kinetic and potential energy, and in Cartesian coordinates for a system of $N$ charged particles it has the form

$$\hat{H} = \hat{K} + \hat{V} = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_i} \nabla_i^2 + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N} \sum_{i<j} q_i q_j \frac{1}{r_{ij}}, \quad (23)$$

where $q_i$ and $q_j$ are the charges of the particles in question and $r_{ij}$ is the distance between them. In equation (23), $m_i$ is the mass of particle $i$ and $\epsilon_0$ is the vacuum permittivity.

For systems larger than two particles, this second order differential equation cannot be solved analytically. We have to use numerical methods to obtain the wavefunctions. In order to make the numerical solution of equation (22) feasible, we have to resort to a number of approximations, the first of which are the Born-Oppenheimer and adiabatic approximations.

### 2.2.1 The Born-Oppenheimer and Adiabatic Approximations

Because the nuclei in the system are around 1000 times more massive than the electrons, it is plausible to assume that the electrons respond instantaneously to any change in the nuclear configuration. Thus our goal is to simplify the computational problem by separating the nuclear and electronic motions from one another. This separation can be introduced in the following manner: We note that in a system of $N_e$ electrons and $N_n$ nuclei, the Hamiltonian of equation (23) can, in a center of mass coordinate system, be represented in the form\(^6\)–\(^7\)

$$\hat{H} = \hat{K}_n + \hat{K}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} + \hat{H}_{mp} = \hat{K}_n + \hat{H}_e + \hat{H}_{mp}, \quad (24)$$

where $\hat{K}_n$ and $\hat{K}_e$ are the nuclear and electronic kinetic energy operators, respectively, $\hat{V}_{ne}$ is the potential energy term between the electrons and the nuclei, $\hat{V}_{ee}$ is the potential energy term between electrons, and $\hat{V}_{nn}$ is
the nuclear potential energy term. In the rightmost equation, the electron dependent terms have been combined to form the electronic Hamiltonian operator $\hat{H}_e$. The mass-polarization operator $\hat{H}_{\text{mp}}$ arises because it is impossible to separate center of mass motion from internal motion in a system of more than two particles.

Because the electronic Hamiltonian operator is Hermitian, the electronic wave functions $\psi_{e,i}(y,x)$ that are solutions to the Schrödinger equation

$$\hat{H}_e \psi_{e,i}(y,x) = E_i(y) \psi_{e,i}(y,x)$$

form a complete orthonormal set of functions. In equation (25), $x$ are the electron coordinates and $y$ are the nuclear coordinates. It is useful to note that both the wave functions and energies depend on the locations of the nuclei. This dependence arises because the nuclear coordinates enter $\hat{H}_e$ as parameters that determine $V_{ne}$ and $V_{nn}$.

The completeness of the $\psi_{e,i}$ set means that it is possible to express the wave functions of the total Hamiltonian as a linear combination of the electronic wave functions:

$$\psi = \sum_{i=1}^{\infty} \psi_{n,i}(y) \psi_{e,i}(y,x),$$

where $\psi_{n,i}(y)$ are the expansion coefficients, which turn out to be the nuclear wave functions.

The nuclear wave functions can be found by operating on this function with the total Hamiltonian of equation (24), multiplying from the right by a specific $\psi_{e,k}$ and integrating over all space. This results in the equation

$$E_{\text{tot}} \psi_{n,k} = \hat{K}_n \psi_{n,k} + E_k \psi_{n,k} + \sum_{i=1}^{\infty} \left( 2 \langle \psi_{e,k} | \nabla_n | \psi_{e,i} \rangle \nabla_n + \langle \psi_{e,k} | \hat{K}_n | \psi_{e,i} \rangle + \langle \psi_{e,k} | \hat{H}_{\text{mp}} | \psi_{e,i} \rangle \right) \psi_{n,i},$$

where $E_i$ is the energy eigenvalue of the electronic Hamiltonian for the state $\psi_{e,i}$, and the operator $\nabla_n$ is defined by equation

$$\nabla^2_n = \hat{K}_n = -\sum_{j=1}^{N_n} \frac{\hbar^2}{2m_{a,j}} \nabla^2_j,$$

where $m_{a,j}$ is the atomic mass associated with the nuclei $j$. Here and henceforth, Dirac bracket notation will be used for the integrals over all space.

In equation (27), the terms under summation represent the coupling between different electronic states. The first two terms in the parenthesis are known as the first and second order non-adiabatic coupling elements. The adiabatic approximation sets these coupling terms equal to zero, leaving only the terms for which $i = k$. It is valid in systems where the reaction takes place on a single potential energy surface and fails, for example, when the reaction contains spin-forbidden transitions, as in many photochemical reactions.\textsuperscript{65,68}

Another approximation that can be made arises from the fact that the mass-polarization operator depends inversely on the total mass of the molecule. As already pointed out, the total mass is much larger than the mass of the electron, which results in the mass polarization term to be negligible in most cases.
With the approximations presented and the fact that the first order non-adiabatic couplings are zero for all except spatially degenerate wave functions, equation (27) gives

$$E_{\text{tot}} \psi_{n,k} = \left( \hat{K}_n + E_k + \langle \psi_{e,k} | \hat{K}_n | \psi_{e,k} \rangle \right) \psi_{n,k}. \quad (29)$$

In the Born-Oppenheimer approximation, the diagonal correction terms $$\langle \psi_{e,k} | \hat{K}_n | \psi_{e,k} \rangle$$ are set to zero. This results in an equation of the form

$$E_{\text{tot}} \psi_{n,k}(y) = \hat{K}_n \psi_{n,k}(y) + E_k(y) \psi_{n,k}(y). \quad (30)$$

According to equation (30), the nuclear motion occurs on a potential energy surface $$E_i(y)$$ that can be obtained by solving the electronic Schrödinger equation.

The Born–Oppenheimer approximation works well for most systems, but fails when two states of the system become energetically very close. The errors resulting from the use of the Born-Oppenheimer approximation are often small and these are largest in systems containing hydrogen nuclei. In the hydrogen fluoride molecule, for example, the Born–Oppenheimer approximation introduces a shift in harmonic wavenumber that is only 1.5 cm$^{-1}$.

Of the two equations (25) and (30), the electronic Schrödinger equation is usually the more difficult one to solve due to the large number of electrons in most chemically interesting systems. Thus, in the coming sections, we will concentrate mainly on the methods developed for solving (25) and for simplicity’s sake, write the electronic Hamiltonian and wave function simply as $$\hat{H}$$ and $$\psi$$, respectively.

The electronic Schrödinger equation retains the general form of (22) with the Born–Oppenheimer approximation, and, for a system consisting of $$N_e$$ electrons and $$N_n$$ nuclei, the electronic Hamiltonian can be written as

$$\hat{H} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{N_e} \nabla_i^2 - \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N_e} \sum_{k=1}^{N_n} \frac{q_k e}{r_{ik}} + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N_e} \sum_{j<k}^{N_e} \frac{e^2}{r_{ij}}, \quad (31)$$

where we have divided the potential energy into two terms one associated with the nuclear–electron attractions and the other with the electron–electron repulsions.

While the Born–Oppenheimer approximation is the fundamental approximation underlying almost all ab initio calculations, for systems containing more than one electron it is impossible to solve equation (31) exactly and further approximations will be required.

### 2.2.2 The Hartree–Fock Approach

The complications in the solution of the electronic Schrödinger equation arise from the electron-electron interaction term in equation (31). In the Hartree–Fock (HF) approach, this interaction is treated in an approximate way, so that each electron travels in the average electric field generated by all the other electrons and the nuclei. The roots of the Hartree–Fock method lie
within the variational theorem, which states that for any trial wave function denoted as \( \psi_t \), the so-called Rayleigh ratio \( \mathcal{E} \) defined as

\[
\mathcal{E} = \frac{\langle \psi_t | \hat{H} | \psi_t \rangle}{\langle \psi_t | \psi_t \rangle}
\]

is always greater than or equal to the system energy \( E \). Thus, the variational theorem states that there exists a lower bound to the energy obtained from our trial wave function, which is always larger than or equal to the true energy.

Given that the electronic wave function has to fulfill the Pauli principle, it is logical to take the trial Hartree–Fock wave function \( \Phi_0 \) in the form of a Slater determinant

\[
\Phi_0 = \frac{1}{\sqrt{N_e!}} \det \left| \varphi_a(1) \varphi_b(2) \varphi_c(3) \ldots \varphi_z(N_e) \right|,
\]

where \( \varphi_a, \varphi_b, \ldots, \varphi_z \) signify the molecular orbitals. Often in equation (33), the one electron wave functions are called spinorbitals. In the absence of an external field, the spinorbitals can be obtained by multiplying the spatial orbital with the spin part of the wave function. Customarily each spinorbital is expanded as a linear combination of a set of \( n \) basis functions \( \xi_p \):

\[
\varphi_u = \sum_{p=1}^{n} c_{pu} \xi_p,
\]

where \( c_{pu} \) are coefficients that we need to determine. The exact nature of the basis functions will be discussed later in section 2.2.6. It is worthwhile to note that from this set of basis functions exactly \( n \) linearly independent wave functions can be formed.

With the electronic Hamiltonian and the trial wavefunction defined by equations (31) and (33), it follows from the variational theorem that the problem in the HF calculations can be viewed as one of minimizing the energy functional \( \langle \Phi_0 | \hat{H} | \Phi_0 \rangle \) under the constraint that the different spinorbitals must remain orthonormal. This can be done by using the method of Lagrange multipliers, and results in a series of Hartree–Fock equations of the type

\[
\hat{f}_i | \varphi_a(i) \rangle = \epsilon_u | \varphi_a(i) \rangle,
\]

where the fock operator is defined as

\[
\hat{f}_i = \hat{h}_i + \sum_{a=1}^{x} \left( \hat{J}_u(i) - \hat{K}_u(i) \right),
\]

where the sum is over the occupied molecular orbitals and \( \epsilon_u \) are orbital energies. In equation (36), \( \hat{h}_i \) is a one electron core Hamiltonian

\[
\hat{h}_i = -\frac{\hbar^2}{2m_e} \nabla_i^2 - \frac{1}{4\pi\epsilon_0} \sum_{j=1}^{N_e} \frac{q_j e}{r_{ij}},
\]

which consists of the kinetic energy of the electron and its interactions with the nuclei. For electrons \( i \) and \( j \) the Coulomb operator \( \hat{J}_u(i) \) is defined by the integral

\[
\hat{J}_u(i) | \varphi_c(i) \rangle = \frac{e^2}{4\pi\epsilon_0} \langle \varphi_u(j) | \frac{1}{r_{ij}} | \varphi_u(j) \rangle | \varphi_c(i) \rangle.
\]
The Coulomb operator accounts for the Coulombic repulsions between electrons. $\hat{K}_u(i)$ is called the exchange operator. It is defined by the integral

$$\hat{K}_u(i) |\varphi_v(i)\rangle = \frac{e^2}{4\pi\epsilon_0} \langle \varphi_u(j) | \frac{1}{r_{ij}} | \varphi_v(j) \rangle |\varphi_u(i)\rangle.$$  \hspace{1cm} (39)

The exchange operator takes into account the spin correlation effect between electrons.

Strictly speaking the spinorbitals appearing in equation (35) are not the same as those appearing in (33), but rather linear combinations called canonical spinorbitals. This causes no actual computational problems, due to the fact that the transformation between the two sets of spinorbitals is unitary and a unitary transform leaves the determinantal total wave function unchanged.

By insertion of equation (34) and multiplication from the left by $\langle \xi_q |$, the Fock equations for all the orbitals are transformed into a matrix equation of the form

$$FC = SCE,$$  \hspace{1cm} (40)

where $C$ is an $n \times n$ matrix of the coefficients and $E$ is an $n \times n$ diagonal matrix of the orbital energies. The elements in the Fock matrix $F$ have the following form

$$F_{qp} = \langle \xi_q(i) | \hat{f}_i | \xi_p(i) \rangle,$$  \hspace{1cm} (41)

and the elements in the overlap matrix $S$ the form

$$S_{qp} = \langle \xi_q(i) | \xi_p(i) \rangle.$$  \hspace{1cm} (42)

Essentially, the problem of determining the best possible determinantal wave function has been transformed into a problem of finding the coefficient matrix. In principle this is easy, because basic linear algebra tells us that the matrix equation (40) can have a non-trivial solution only if the condition

$$\det |F - \epsilon_u S| = 0$$  \hspace{1cm} (43)

is satisfied. This equation is dealt with by giving an initial set of coefficients and then calculating the Fock and overlap matrices from equations (41) and (42), respectively. From these a new set of orbital energies and coefficients are obtained, which can then be used to recalculate the Fock and overlap matrices and so on, forming an iterative process. Usually the cycle is repeated until the difference between the new and old coefficients is negligible and the system has achieved self-consistency. Regardless of the choice of basis functions, the variational theorem in equation (32) guarantees that the final energies are always greater than the true energy of the system.

### 2.2.3 Electron Correlation

It has been approximated that the HF-method accounts for about 99% of the total energy. The remaining one percent arises from the instantaneous coulombic and spin correlation effects, which cause the electrons to avoid each other more than the average HF treatment would propose. The difference between the HF energy and the lowest possible energy for a given basis function set is called the electron correlation energy. Unfortunately, this one percent is essential when we are dealing with chemical phenomenon.
Electron correlation can be classified in a few ways. First, it can be divided between static and dynamic correlation, where static correlation represents the more permanent kind of correlation between electrons that are far away from each other, such as on different spatial orbitals. Dynamic correlation is the instantaneous correlation between two electrons that occupy the same spatial orbital. An alternative distinction is between Fermi and Coulomb correlation, where Fermi correlation is the correlation between the same spin electrons and Coulomb correlation is between electrons of the opposite spin.

Because the HF approach gives the best one-determinantal wave function \( \Phi^0 \) as its solution for the ground state wave function, additional determinants have to be considered if electron correlation is to be taken into account. These determinants can be constructed from the leftover \( n - N_e \) virtual orbitals that result from filling the lowest of our \( n \) spinorbitals with \( N_e \) electrons. The different types of excited determinants \( \Phi_1, \Phi_2, \ldots \) are formed by promoting electrons to the virtual spinorbitals \( \varphi_\alpha \). For example, in the case of two electron promotion from spinorbitals \( b \) and \( c \) in equation (33) to the virtual orbitals \( \varphi_\alpha \) and \( \varphi_\beta \), we would have one of the doubly excited determinants:

\[
\Phi_{bc}^{\alpha\beta} = \frac{1}{\sqrt{N_e!}} \det |\varphi_a \varphi_\alpha \varphi_\beta \ldots \varphi_z|.
\]

(44)

These determinants, or small linear combinations of them constructed to achieve the correct electron symmetry, are eigenfunctions of all the operators that commute with \( \hat{H} \) and are called configuration state functions (CSFs). Depending on the electron correlation method used, the actual wave function is then represented as a some kind of linear combination of the different CSFs.

The three most common ways to deal with electron correlation are Configuration Integration, Møller-Plesset Many Body Perturbation (MP), and Coupled Cluster (CC) methods. In this section only MP and CC methods will be explored, as they are the ones used in this study.

2.2.4 MP2

The Møller–Plesset approach\(^{75}\) applies many-body perturbation theory to the electron correlation problem. Perturbation theory makes use of the fact that for some small perturbations the Hamiltonian, the ground state wave function and the ground state energy can be expanded as

\[
\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)} + \hat{H}^{(2)} + \ldots
\]

(45)

\[
E = E^{(0)} + E^{(1)} + E^{(2)} + \ldots
\]

(46)

\[
\psi = \psi^{(0)} + \psi^{(1)} + \psi^{(2)} + \ldots
\]

(47)

Here, \( \hat{H}^{(0)} \) is some easily manageable Hamiltonian that is a relatively good approximation of the ground state of the system. In the MP method, it is chosen to be the sum of the Fock operators encountered in equation (36)

\[
\hat{H}^{(0)} = \sum_{i=1}^{N_e} \hat{f}_i.
\]

(48)

As can be seen from the definition of the fock operator (36), this leads to double counting of the electron electron repulsion. To remedy this, the first
order correction $\hat{H}^{(1)}$ is defined by

$$\hat{H}^{(1)} = \hat{V}_{ee} - \sum_{i=1}^{N_e} \sum_{u=a} \left( \hat{J}_u(i) - \hat{K}_u(i) \right),$$

(49)

and the higher order perturbation Hamiltonians are set equal to zero, i.e., $\hat{H}^{(2)} = \hat{H}^{(3)} = \cdots = 0$. As in equation (24), $\hat{V}_{ee}$ is the electron-electron repulsion operator.

The perturbation wave functions of equation (47) are represented as a linear combination of the excited Slater determinants $\Phi_J$. For example, the first order correction to the wavefunction is of the form

$$\psi_0^{(1)} = \sum_{J \neq 0} C_J \Phi_J,$$

(50)

where the sum is over all excited determinants, and the coefficients can be found from

$$C_J = \frac{\langle \Phi_J | \hat{H}^{(1)} | \Phi_0 \rangle}{E^{(0)} - E_J^{(0)}},$$

(51)

where $E_J^{(0)}$ is the energy eigenvalue of the $\Phi_J$ eigenfunction of the zeroth-order Hamiltonian and $\Phi_0$ is the ground state Hartree–Fock wave function. We observe from equations (50) and (51) that perturbing the system results in the wave function being a mixture of the ground and excited states. From the denominator, we see that the contribution of a given state decreases with increasing energy of the state, so that the closest lying states tend to have the largest impact on the wave function. Because the electronic Hamiltonian operator only consists of terms that depend on one or two electronic coordinates, it follows from the orthonormality of the $\Phi$ states that all matrix elements containing more than doubly excited determinants vanish. On the other hand, Brillouin’s theorem states the matrix elements containing singly excited determinants become zero, and the sum is thus only over the different doubly excited determinants.

From the energy perspective, first order perturbation has to be taken into account just to reach the HF energy because of the overcounting effect. This means that the first improvement upon the HF energy results from second order perturbation, where the energy correction is of the form

$$E^{(2)} = \sum_{J \neq 0} C_J \langle \Phi_J | \hat{H}^{(1)} | \Phi_J \rangle = \sum_{J \neq 0} \frac{\langle \Phi_J | \hat{H}^{(1)} | \Phi_0 \rangle \langle \Phi_0 | \hat{H}^{(1)} | \Phi_J \rangle}{E^{(0)} - E_J^{(0)}},$$

(52)

By the application of Brillouin’s theorem and Slater-Condon rules this expression can be transformed to a simple sum of two electron integrals

$$E^{(2)} = \sum_{\text{occ}} \sum_{\text{vir}} \frac{\left( \varphi_\alpha \varphi_u | \varphi_\beta \varphi_v \right) - \left( \varphi_\beta \varphi_u | \varphi_\alpha \varphi_v \right)}{\epsilon_u + \epsilon_v - (\epsilon_\alpha + \epsilon_\beta)},$$

(53)

where $\epsilon$ are the spinorbital energies, and the sums go over all occupied and virtual orbitals, respectively. The total energy is obtained simply by adding the perturbation energy of equation (53) to the ground state energy as shown in equation (46). Equation (53) used the Mulliken notation (for some arbitrary electrons $i$ and $j$)

$$\left( \varphi_\alpha \varphi_u | \varphi_\beta \varphi_v \right) = \frac{e^2}{4\pi \epsilon_0} \frac{1}{r_{ij}} \langle \varphi_u(i) | \varphi_v(j) \rangle.$$
\[
e^2 \frac{1}{4\pi \epsilon_0} \langle \phi_u(i)\phi_v(j) | \frac{1}{r_{ij}} | \phi_\alpha(i)\phi_\beta(j) \rangle. \tag{54}
\]

As can be seen in equation (52), the determination of the second order perturbation energy requires no knowledge of the second-order wave function. It can be shown that the knowledge of the \(m\)th order wave function allows one to calculate the perturbed energy up to order \(2m + 1\). This means that the second order correction to the wave function is only required to achieve energy corrections of fourth and fifth orders.

The method that employs only the second order correction to the energy is called MP2. Physically, the second order perturbation accounts for interactions between the two electrons, which form the largest contribution into the electron correlation energy. It has been reported that MP2 recovers roughly 80 – 90% of the electron correlation energy. MP3 introduces third order perturbation in energy which accounts for interactions between electron pairs and allows for the orbitals to relax in some sense from the second order picture. However, this effect sometimes leads to underestimation of the true energy, and so MP3 might perform poorer than MP2, which often overestimates the true energy. The fourth order correction in MP4 introduces single, triple and quadruple excitations to the perturbation energy, and is able to account for approximately 95 – 98% of the correlation energy.

Computationally, an MP2 calculation consists of three parts: An HF calculation to obtain the reference Slater determinant which has to describe the true system well enough for the perturbation calculations to converge. In MP2, the computational effort of this phase scales as \(n^4\) where \(n\) is the size of the basis set. The bottleneck of the calculation is the second phase where the transformation of the atomic orbitals or basis functions of equation (34) to the molecular orbital basis takes place. In MP2, this phase scales as \(n^5\). Finally, the MP2 calculation for the energy is relatively easy to do and scales as \(n^4\). In MP3, the computational effort of the second phase increases to \(n^6\) and finally to \(n^7\) in MP4.

The convergence behavior of the different MP\(m\) methods varies from one system to another and depends on whether the electron pairs of the system are well separated or cluster together in some regions. In systems with separated electron pairs, the convergence is usually monotonous and pair correlation effects dominate the correlation energy. On the other hand, clustered systems show oscillating convergence behavior, and correlation arises chiefly from three electron interactions and pair correlation effects in the areas of high electron density. Because the convergence properties of the MP\(m\) series vary greatly with the number of diffuse functions within the basis set used and generally diverges for diffuse enough basis, the use of higher order MP-calculations can be called into question especially in systems where a diffuse basis has to be used for chemical reasons.

Important concepts regarding electron correlation methods are size consistency and size extensivity. For size consistent methods, the energy for the simultaneous calculation of two non-interacting systems \(A\) and \(B\) equals the sum of the individual energies of the systems. This means that the energy of the system scales properly with the number of electrons. In a size extensive system, the energy scales properly with the number of particles in the system, meaning that all the particles in the system can be allowed to interact. All the MP\(m\) methods are both size extensive and size consistent,
but because the perturbational corrections to the ground state energy can be positive or negative depending on the perturbation, the MP-methods are not variational. It is due to this lack of iterative processes that the MPm methods are in general about an order of magnitude more efficient than the corresponding Configuration Interaction or CC methods.\textsuperscript{20}

**Local Approximations**

The canonical orbitals obtained from a HF calculation are, by their nature, delocalized over the whole molecule. As far as the physical interpretation is concerned, however, electron correlation effects arise chiefly from electrons that are near one another. Since the delocalization of the canonical orbitals is the reason for the sharp, non-physical scaling of the standard electron correlation methods, it seems reasonable that its effects could be reduced by doing calculations in a localized set of orbitals.\textsuperscript{79}

Mathematically, the locality of the correlation effects can be seen from the fact that the ground HF reference wave function $\Phi_0$ is connected to the doubly substituted determinants $\Phi_2$ by the matrix element $(\varphi_u \varphi_\alpha | \varphi_v \varphi_\beta)$ according to equation (51). It follows that the only significant contributions to the wave function arise from virtual orbitals for which the differential overlap terms $\varphi_u \varphi_\alpha$ and $\varphi_v \varphi_\beta$ are large, i.e., from virtual orbitals which are localized within the same region of space as the original occupied orbitals.

Local correlation methods simplify calculations by omitting those two electron integrals where the differential overlaps are small. This is done separately for each electron by restricting the virtual space to the subspace of the atomic orbitals in the neighbourhood of the orbitals to be correlated. In the successful formulation by Pulay and Sæbo,\textsuperscript{20,80–82} each of the correlated occupied orbitals $|\varphi_u\rangle$ is assigned a domain $D(u)$ that is independent of molecular size. For a bond with $n$ centers, the $D(u)$ consists of valence atomic orbitals (AOs) of the $n$ atoms on which it is localized, projected onto the virtual orbital basis to ensure orthogonality. The local basis for a pair of occupied orbitals $|\varphi_u\rangle$ and $|\varphi_v\rangle$ is defined as $D(u) \cup D(v)$. Ideally, the virtual basis is chosen so that it involves all pair correlations, leads to smooth potential energy surfaces and is still small enough to cause computational savings.

There exist a few algorithms which are commonly in use for the localization of the orbitals. The least expensive one is the Boys method\textsuperscript{83} which minimizes the sums of squared orbital radii. This method tends to have difficulty in radial localization.\textsuperscript{20} Alternatively, one can use either the Edmiston-Ruedenberg\textsuperscript{84} or Pipek–Mezey\textsuperscript{85} methods, both of which maximize intraorbital Coulomb repulsion. Because the total determinantal wave function is invariant in all nonsingular transformations of the molecular orbitals, it is possible to choose the spatially localized functions so that orthogonality is retained. Localization is usually carried out only for the valence orbitals due to the naturally localized structure of the core orbitals.\textsuperscript{20}

Compared with orbitals that are on the same atom, the correlation energies of orbitals that are separated by one bond are approximately an order of magnitude smaller, and for weaker pairs the absolute pair energies decrease as $r^{-6}$. Thus, in addition to the truncation of the virtual space, further savings can be obtained by imposing a hierarchy upon the electron pairs.
within the molecule. The modern version of the localized treatment by Werner et al.\textsuperscript{21,22,86} divides the electron pairs into four groups according to the minimum distance $R$ between any atoms within the domains $D(u)$ and $D(v)$. In their treatment, strong pairs have $R \leq 1$ bohr, and the pairs are treated at the Coupled Cluster singles and doubles level—a variant of the Coupled Cluster methods described in section 2.2.5. Weak pairs with $1 < R \leq 8$ bohr are treated by the local MP2. For distant pairs with $8 < R \leq 15$ bohr, long range ionic excitations are neglected and a multipole expansion is used.\textsuperscript{87} Finally, very distant pairs, for which $R > 15$ bohr, are completely neglected. Because the number of strong, weak and distant pairs all scale linearly with the system size, larger computational saving can be obtained with this partitioning than with the original one by Pulay and Saebø.\textsuperscript{21,82}

The localized orbitals are no longer eigenfunctions of the Fock operator of equation (36), and so an iterative process must be used for the calculation of the perturbation energy.\textsuperscript{21,80} In spite of this, the slowest part of this calculation is the transformation of the two electron integrals of equation (54) into the atomic orbital basis as in the standard MP2. To bring down the scaling of this step, Werner et al. made use of efficient screening techniques and split the coulomb operator in (54) into a rapidly decaying short range and a smooth long range part and then used a multipole expansion to efficiently calculate the latter.\textsuperscript{21,22}

In addition to the radical reduction in computational cost from $n^5$ to linear dependency with the size of the system,\textsuperscript{22} the local correlation scheme has other advantages compared with the ordinary MP2. Perhaps the most important of these is that the truncation of the virtual space greatly reduces both the intermolecular and intramolecular basis set superposition errors (BSSEs),\textsuperscript{88} making the counterpoise (CP) correction\textsuperscript{89} unnecessary. Finally, the local correlation method produces geometries of comparable accuracy and slightly improved vibrational frequencies for a fraction of the computational cost of the regular methods.\textsuperscript{90–92}

The BSSE arises from the fact that the quality of a truncated nuclear centered basis set is not equal for different geometries. For example in a complex such as the sulfuric acid monohydrate, the basis functions on one of the different monomers can help to compensate for the incompleteness of the basis set in the other. This leads to an artificial lowering in the energy of the complex with respect to the monomer energies and ultimately to an overestimation of the hydrogen bond strength. The CP correction is an approximate method to account for the BSSE.

### Density Fitting

Even though the local approximations can be used to achieve linear scaling with molecular size, they do nothing to the steep fourth-order scaling of the computational cost with basis set size per atom.\textsuperscript{91} To reduce the scaling, density fitting (DF) is a procedure that creates savings in the computational time required for the four electron integrals by making use of the resolution of identity:\textsuperscript{72,93}

$$1 = \sum_U |\xi_U\rangle \langle \xi_U|,$$

(55)
where the $\xi_U$ functions form a complete orthonormal set. Insertion of this into equation (54) gives

$$
(\varphi_\alpha \varphi_u | \varphi_\beta \varphi_v) = \sum_U \langle \varphi_u | \varphi_\alpha \xi_U \rangle \langle \xi_U | \varphi_\beta \varphi_v \rangle.
$$

The first integral on the right is a three-index overlap integral and the second integral is defined analogously to equation (54) with the $r_i$ dependent functions replaced by the single $\xi_U(r_i)$. In equation (56), the problematic four electron integral has been replaced by a sum over the products of two three electron integrals. These integrals are easier to compute and store than the original ones, so in principle substantial savings in computation time could be achieved.

However, because it is impossible in practice to utilize a complete basis set of the auxiliary $\xi_U$, the actual method used in density fitting approaches differs slightly from the straightforward application of the resolution of identity described above. Following the approach of Werner et al., we begin by noting that the products of the two molecular orbitals $\varphi_\alpha \varphi_u = \rho_{\alpha u}$ are actually electron densities for the single electron. Now we use the auxiliary basis set to approximate $\rho_{\alpha u}$ by the formula

$$
\tilde{\rho}_{\alpha u} = \sum_{\text{fit}} d_{\alpha u}^{\text{fit}} \xi_U.
$$

(57)

Usually, the functions $\xi_U$ are chosen to be the atom-centered Gaussian type orbitals described in section 2.2.6. The coefficients $d_{\alpha u}^{\text{fit}}$ are calculated by minimizing the Coulomb energy fitting residual

$$
\epsilon_{ab} = \int \int [\rho_{ab}(r_i) - \tilde{\rho}_{ab}(r_i)][\rho_{ab}(r_j) - \tilde{\rho}_{ab}(r_j)] \rho_{ij} d\mathbf{r}_i d\mathbf{r}_j,
$$

(58)

which also minimizes the least squares error of the electric field. This procedure results in the equations

$$
d_{\alpha u}^{\text{fit}} = \sum_{\text{fit}} \langle \varphi_\alpha \varphi_u | \xi_U \rangle [J^{-1}]_{UV}.
$$

(59)

With this, the four electron sum is approximated by the term

$$
(\varphi_\alpha \varphi_u | \varphi_\beta \varphi_v) = \sum_{\text{fit}} d_{\alpha u}^{\text{fit}} \langle \xi_U | \varphi_\beta \varphi_v \rangle
$$

$$
= \sum_U \langle \varphi_\alpha \varphi_u | \xi_U \rangle [J^{-1}]_{UV} \langle \xi_U | \varphi_\beta \varphi_v \rangle,
$$

(60)

where $J$ is an $n_{\text{fit}} \times n_{\text{fit}}$ matrix with the elements

$$
J_{UV} = \int \int \xi_U(r_i) \xi_V(r_j) \rho_{ij} d\mathbf{r}_i d\mathbf{r}_j.
$$

(61)

The DF-MP2 calculations proceed in four stages: First the minimization procedure of equation (58) is used to calculate the expansion coefficients. Then the transformation process from the atomic orbital basis to the molecular orbital basis is done in two separate stages:

$$
\sum_q c_{qu} \langle \xi_p \xi_q | \xi_U \rangle = \langle \xi_p \varphi_u | \xi_U \rangle
$$

(62)
\[ \sum_p c_{pa}(\xi_p \varphi_u | \xi_U) = (\varphi_u \varphi_u | \xi_U). \] (63)

After this transformation, the linear equations of (59) are solved and finally equation (60) is used to calculate the two electron integrals and the MP2 energy. Considering these four stages, the evaluation of equation (60) is computationally the most expensive and formally scales as \( n^5 \).\(^{23}\)

In the above formulation of the density fitting method, a canonical set of orbitals is assumed. A local treatment can be used to decrease the computational effort of all the relevant steps. The bottleneck computational effort of equation (60) can be reduced to \( n^2 \). Linear scaling is achieved when the localization scheme is extended to the fitting functions by introducing local fitting domains.\(^{23,101}\) This means that for each electron, a different fitting basis is used, and the linear equation (59) is solved separately in each case. The fitting domains for electron pairs are formed by combining all fitting functions at the atoms in the pair domain, and adding all fitting functions on atoms at most some distance \( R_d \) away from the atoms in the domain.

With the right choice of the auxiliary basis set, the error arising from the density fitting approximation becomes negligible\(^{102}\) and third order scaling with basis set size per atom is achieved.\(^{23}\) The density fitting and local approximations can also be applied to the HF energy calculation enabling calculations of large systems with more than 4000 basis functions.\(^{103}\)

**Spin Component Scaling**

It is well-known that the starting point of the canonical MP series contains a biased description of the electron pairs with different and same electron spins. This arises because at the HF level the Fermi correlation between spin-parallel pairs is taken into account, while the Coulomb correlation of spin-anti-parallel pairs is neglected.\(^{24}\) At the MP2 level, this bias is continued, which leads to a systematic overestimation of the static electron correlation energy and an underestimation of the dynamic electron correlation.

The spin component scaling (SCS) method originally proposed by Grimme\(^{25}\) offers a simple way of accounting for this bias. Because the Hamiltonian in (31) is independent of spin, it is possible to separate the MP2 correlation energy into a sum of terms arising from parallel spin pairs and anti-parallel spin pairs.\(^{104}\) In the SCS approximation, the unequal errors in these two energies arising from the biased treatment of MP2 are corrected by introducing separate scaling of the energies in the form of parameters \( p_T \) and \( p_S \):

\[ E^{(2)} = p_T E_T^{(2)} + p_S E_S^{(2)}, \] (64)

where \( E_T^{(2)} \) is the MP2 correlation energy for electrons with parallel spin pairs and \( E_S^{(2)} \) stands for the MP2 correlation energy for electrons with antiparallel spin pairs. In an SCS calculation, these energies are obtained separately and then scaled according to (64) to give \( E^{(2)} \).

The SCS method improves the accuracy of the original MP2 method without any noteworthy increase in computational cost.\(^{24}\) On the average, the SCS-MP2 recovers somewhat more correlation energy than MP2, but more importantly it creates a more uniform distribution of results than MP2: With the coefficient values \( p_S = \frac{5}{6} \) and \( p_T = \frac{1}{3} \) there is a significant drop in
the root mean square (RMS) deviations with respect to experimental values for bond lengths and vibrational frequencies, and in the RMS values with respect to a QCISD(T)/QZV(3d2f,2p1d) calculation for the reaction energies. In addition to the original scaling coefficients quoted above, other authors have suggested various values for the coefficients. It has been noted that the optimal scaling parameters depend on the set of basis functions employed, the chemical nature of the problem and the properties of interest.

It is possible to combine all the additional approximations discussed so far. This DF-SCS-LMP2 method inherits the good qualities of its component methods, meaning that it is both accurate, fast and computationally inexpensive. It has been shown, for example, that in a π-stacked benzene dimer system, the DF-SCS-LMP2 with an aug-cc-pVTZ* basis set is capable of reproducing the approximate Coupled Cluster calculations at the CCSD(T) level with a quintuple zeta aug-cc-pVQZ* basis set. Both basis sets are identical to the aug-cc-pV\textsubscript{m}Z ones, where \( m \) is T or Q, except that only the correlation consistent cc-pV\textsubscript{m}Z basis has been used for the hydrogens.

### 2.2.5 Coupled Cluster Methods

Whereas in the MP\textsuperscript{m} methods all types of electron excitations in equation (44) are included to a certain order \( (m) \), the Coupled Cluster methods incorporate all orders of electron excitations up to a given type. This is done with the help of the the cluster operator \( \hat{T} \) defined by equation

\[
\hat{T} = \hat{T}_1 + \hat{T}_2 + \ldots + \hat{T}_{N_e}, \tag{65}
\]

where \( \hat{T}_n \) are excitation operators. When operating on the HF wave function, the \( \hat{T}_n \) generate a set of excited determinants of a given order \( n \). For example \( \hat{T}_2 \) operating on the HF reference wave function results in a series of determinants of the form

\[
\hat{T}_2 \Phi_0 = \sum_{u<v} \sum_{\alpha<\beta} t_{uv}^{\alpha\beta} \Phi_{\alpha\beta}^{uv}, \tag{66}
\]

where the coefficients \( t \) are called amplitudes. With these notations the Coupled Cluster wave function is defined as

\[
\psi_{\text{CC}} = e^{\hat{T}} \Phi_0 = \left( 1 + \hat{T} + \frac{1}{2!} \hat{T}^2 + \frac{1}{3!} \hat{T}^3 + \ldots \right) \Phi_0
\]

\[
= \left[ 1 + \hat{T}_1 + \left( \hat{T}_2 + \frac{1}{2} \hat{T}_1^2 \right) + \left( \hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{3!} \hat{T}_1^3 \right) + \ldots \right] \Phi_0. \tag{67}
\]

Each term in parenthesis is responsible for incorporating all orders of a given type of excitation (single, double, etc.). The physical difference between the different terms of a given order can be explained by looking at, for example the triple excitation terms \( \hat{T}_3 \) and \( \hat{T}_2 \hat{T}_1 \). In \( \hat{T}_3 \), the excitations are connected, meaning that they interact with each other, while the disconnected term \( \hat{T}_2 \hat{T}_1 \) describes the interaction between a single electron and a pair. Because it becomes increasingly unlikely to have a large number of electrons interacting at the same time, it can be assumed that the importance
of the $\hat{T}_k$ terms decrease as $k$ increases. As remarked in section 2.2.4, the most important interaction arises from the $\hat{T}_2$ term when a canonical set of molecular orbitals is in use. As a result, the quadruply excited $\hat{T}_2^2$ is likely to be the most important term in the quadruple excitations.

When the whole Taylor expansion up to $\hat{T}_{N_e}$ is used, all correlation energy for the given basis is obtained. However, for systems of any size, this is completely unfeasible and the series has to be truncated at some point. Due to Brillouin’s theorem, the Coupled Cluster method that involves only single excitations (CCS) would yield the HF energy. This means that the singles only influence the correlation energy through their interactions with the excited configurations, and so their total effect on the correlation energy is small. Thus, the lowest level of improvement follows by treating only the double excitations (CCD), which account for the two electron excitations. Further adding the singles (CCSD) has little impact on the computational effort and slightly improves the results by introducing orbital relaxation. 65

There are in principle two different ways in which the CC energy can be calculated from the Hamiltonian equation

$$\hat{H}e^{\hat{T}}\Phi_0 = E_{\text{CC}}e^{\hat{T}}\Phi_0. \quad (68)$$

First would be to use the variational theorem (32) by choosing $\psi_t = e^{\hat{T}}\Phi_0$. Solutions of this type have clear benefits, as in addition to being variational they can be used for systems with strong electron correlations where the second method fails. Unfortunately because equation (67) displays factorial complexity even for truncated $\hat{T}$, the variational solution is only feasible in small systems.

In the second method, the Coupled Cluster Schrödinger equation is projected to the reference wave function $\Phi_0$ by multiplying the left hand side of equation (68) with $\langle \Phi_0 |$. This yields the correlation energy expression

$$E_{\text{CC}} = \langle \Phi_0 | \hat{H}e^{\hat{T}} | \Phi_0 \rangle = E_0 + \sum_{\text{occ}} \sum_{\text{vir}} \sum_{\alpha<\beta} \left( t_{\alpha\beta}^{\alpha\beta} + t_{\alpha\beta}^{\alpha} t_{\alpha\beta}^{\beta} - t_{\alpha\beta}^{\beta} t_{\alpha\beta}^{\alpha} \right) \left( \langle \varphi_\alpha | \varphi_\beta | \varphi_\alpha \varphi_\beta \rangle \right). \quad (69)$$

As can be seen from equation (69), the energy depends only on two-electron molecular orbital integrals and the singles and doubles amplitudes. The unknown amplitudes can be found by projecting the Schrödinger equation onto space spanned by the excited determinants. Usually this is done via a similarity transform of the Hamiltonian, which can be obtained from equation (68) by multiplication from the left with $e^{-\hat{T}}$:

$$e^{-\hat{T}} \hat{H}e^{\hat{T}} \Phi_0 = E_{\text{CC}} \Phi_0. \quad (70)$$

From this, a set of algebraic equations for the coefficients can be obtained by multiplication from the left with $\langle \Phi_0^a |$, $\langle \Phi_0^{a\beta} |$ etc. These equations can be solved by iterative techniques and from the coefficients the energy in (68) can be calculated. 65,110

Because the CCSD method scales as $n^6$, and the addition of higher excitations causes an increase in computational effort that is two orders of magnitude per level, CCSD is usually the only affordable pure CC method for large
systems. However, because the effects of the $\hat{T}_1$ terms are so small, most of the effect from the triple excitations arise from the $\hat{T}_3$-term in (67). This term can be implemented into standard CCSD by using perturbation theory. In the most effective formulation of the different perturbational approaches, denoted CCSD(T) and introduced by Raghavachari et al.,\textsuperscript{111} the $\hat{T}_3$-term is evaluated by an MP4 calculation with the original CCSD amplitudes and then added to the final energy.\textsuperscript{112} This process increases the computational scaling to $n^7$, but also markedly increases the amount of correlation energy obtained.\textsuperscript{65}

Compared with the MP2 methods, CCSD and CCSD(T) give substantially more accurate results, but at a higher cost. The better performance can be expected because of the low importance of the highly excited $\hat{T}_k$ terms with increasing $k$.

**CCSD(T)-F12 and F12a Methods**

The problematic electron-electron interaction terms in equation (31) cause singularities in the potential energy, which introduce cusps to the wave function at points where two of the electrons coincide. Because the kinetic energy has to cancel the infinity of the potential energy this results in the wave function behaving linearly around the points $r_{ij}$.\textsuperscript{65} Most of the convergence problems in the standard methods arise because the slowly varying orbital product form $\varphi_u(r_i)\varphi_v(r_j)$ is ill-suited for describing the cusp when the electrons $i$ and $j$ are near each other.

To implement the correct cusp behaviour, the CCSD(T)-F12 and CCSD(T)-F12a methods by Adler et al.\textsuperscript{18} define the wave function by

$$|\psi_{\text{CCSD(T)-F12}}\rangle = \exp \left( \hat{T}_1 + \hat{T}_2 \right) |\Phi_0\rangle.$$  \hspace{1cm} (71)

Making use of the Einstein summation rule, where summation over indexes not present on both side of the equation is implied, the cluster operators $\hat{T}_1$ and $\hat{T}_2$ are defined by equations

$$\hat{T}_1 = t_u^\alpha \hat{C}_u^\alpha$$  \hspace{1cm} (72)

$$\hat{T}_2 = \frac{1}{2} T_{uv}^{\alpha\beta} \hat{C}_{uv}^{\alpha\beta} + \frac{1}{2} T_{RS}^{uv} \hat{C}_R^u \hat{C}_S^v,$$  \hspace{1cm} (73)

where $\hat{C}_u^R$, $\hat{C}_{uv}^{\alpha\beta}$ are one and two electron excitation operators, the first of which defines an excitation into a formally complete virtual space $\{R,S,\ldots\}$. It follows from the completeness of this space that it can be partitioned into the union $\{R,S,\ldots\} = \{\alpha,\beta,\ldots\} \cup \{X,Y,\ldots\}$ where $\{X,Y,\ldots\}$ is a complementary auxiliary basis set. Equation (72) and the first term on the right in equation (73) represent the standard first and second order excitations to virtual orbitals as in equation (67).

The added amplitudes $T_{uv}^{\alpha\beta}$ can be found from equations

$$T_{RS}^{uv} = T_{mn}^{uv} F_{RS}^{mn}$$  \hspace{1cm} (74)

$$F_{RS}^{mn} = \langle \varphi_m \varphi_n | F_{12} Q_{12} | \varphi_R \varphi_S \rangle,$$  \hspace{1cm} (75)

where $T_{mn}^{uv}$ are the amplitudes used in the F12 treatment and $m$ and $n$ refer to some occupied orbitals. $F_{12}$ is the short-range correlation factor between
electrons one and two. It has been found that the most intuitive choice of setting $F_{12} = r_{12}$, as done originally by Kutzelnigg and Klopper,\textsuperscript{14,113,114} is not the optimal one when using relatively small basis sets of double zeta or triple zeta quality.\textsuperscript{16} It was shown by Tew and Klopper\textsuperscript{15} that the best correlation factor comes in the form of a Slater function, which can be further fitted to a set of Gaussian geminals\textsuperscript{115} to make the computations easier:

$$F_{12} = \frac{1}{\gamma} e^{-\gamma r_{12}} \approx \sum_k c_k e^{-a_k r_{12}^2},$$

where the coefficients $c_k$ and $a_k$ are determined by least squares fitting\textsuperscript{116} and $\gamma$ is a parameter.

The projector operator $\hat{Q}_{12}$ in equation (75) ensures that the different $F_{12}$ configurations

$$|\Phi_{mn}^{RS}\rangle = F_{mn}^{RS} \hat{C}_R \hat{C}_S |\Phi_0\rangle$$

are orthogonal with the configurations obtained from the standard MO basis. It has the form

$$\hat{Q}_{12} = (1 - \hat{o}_1) (1 - \hat{o}_2) (1 - \hat{v}_1 \hat{v}_2),$$

where $\hat{o}_i = |\psi_u(i)\rangle \langle \psi_u(i)|$ projects on the occupied subspace and $\hat{v}_i = |\psi_v(i)\rangle \langle \psi_v(i)|$ projects on the virtual subspace. In light of definitions (74), (75), (77), (78), the terms $F_{mn}^{RS}$ can be thought of as contraction coefficients between the larger set of configurations $|\Phi_{RS}\rangle$ and the smaller $|\Phi_{mn}\rangle$ one.

In summary, the effect of the additional amplitudes is to introduce new functions into the conventional CC expansion where the products $|\psi_u(1)\psi_v(2)\rangle$ have been replaced with a negative short-range correlation function

$$|\chi_{uv}(1,2)\rangle = T_{mn}^{uv} \hat{Q}_{12} F_{12} |\psi_m(1)\psi_n(2)\rangle.$$  \hspace{1cm} (79)

This both decreases the chance of finding two electrons in the same place, and improves the behaviour of the wave function in the vicinity of the cusp. The most important terms in (79) are those for which $mn = uv$ or $mn = vu$. It is possible to simplify calculations by ignoring less important terms in the wave functions $T_{mn}^{uv} |\Phi_{mn}^{uv}\rangle$.\textsuperscript{18} In this fixed amplitude scheme, the amplitudes $T_{mn}^{uv}$ in the wave function are set to zero for all terms except $T_{uu}^{uu}$, $T_{uv}^{uv}$ and $T_{vu}^{vu}$, which are given fixed values so that the wave function fulfills the cusp conditions. The advantages of this kind of fixed amplitude ansatz is that it is unitary invariant, free of the geminal basis set superposition error and size consistent.\textsuperscript{117}

In a similar manner with the conventional CCSD, equations the CCSD-F12 energy and amplitudes are found by a multiplication from the left with a suitable configuration. Exactly as in CCSD, energy is obtained by the multiplication of $|\Phi_0\rangle$. As for the amplitudes, multiplication by the contravariant configurations

$$|\hat{\Phi}_u^\alpha\rangle = \frac{1}{2} \hat{C}_u^\alpha |\Phi_0\rangle$$

$$|\hat{\Phi}_{uv}^{\alpha\beta}\rangle = \frac{1}{6} \left( 2 \hat{C}_{uv}^{\alpha\beta} - \hat{C}_{vu}^{\alpha\beta} \right) |\Phi_0\rangle$$

results in equations for the singles and doubles residuals

$$R_{\alpha}^u = \langle \hat{\Phi}_u^\alpha | \hat{H} - E | \psi_{\text{CCSD(T)}-F12}\rangle.$$  \hspace{1cm} (82)
\[ R_{\alpha\beta}^{uv} = \left\langle \hat{\phi}_{uv}^{\alpha\beta} \hat{H} - E \right| \psi_{\text{CCSD(T)-F12}} \right\rangle. \]  \hfill (83)

The best values for the coefficients are then obtained by requiring that both equations (82) and (83) equal zero. Because of the fixed \( T_{mn}^{uv} \) amplitudes, the number of equations is exactly the same as in the standard CCSD formulation. The equations themselves differ, due to the additional terms arising from the explicitly correlated terms. Some of these new terms can be simplified by the use of the resolution of identity.\(^{18}\)

The more difficult doubles residual of equation (83) can be represented in matrix form in the basis of the virtual orbitals \((\alpha, \beta)\) as\(^{18,118}\)

\[ R_{\text{CCSD-F12}}^{uv} = R_{\text{MP2-F12}}^{uv} + K(D^{uv}) + K(T^{uv}) + \alpha_{uv,kl}D^{kl} + G^{uv} + G^{uv\dagger}, \]  \hfill (84)

where \(k\) and \(l\) run over occupied orbitals. The first term on the right is the MP2-F12 residual, and is given by the equation\(^{116}\)

\[ R_{\text{MP2-F12}}^{uv} = K^{uv} + FT^{uv} + T^{uv}F - F_{uk}T^{kv} - T^{uk}F_{kv} + C^{mn}T_{mn}^{uv}, \]  \hfill (85)

where \(T^{uv}\) are amplitude matrices, the terms \(F_{rs} = \langle \varphi_r | \hat{f} | \varphi_s \rangle\) are blocks of the closed shell Fock matrix, with the Fock operator \(\hat{f}\) defined by equation (36) and \(K^{uv}_{\alpha\beta} = \langle \varphi_{\alpha} \varphi_{u} | \varphi_{\beta} \varphi_{v} \rangle\) are the usual exchange integrals. The \(r\) and \(s\) in the fock matrix element refer to any orbitals representable in the AO basis. All the irregularities arising from the explicitly correlated terms are incorporated in the last term of equation (85) which describes the coupling to the explicitly correlated configurations. In the F12 method in question, it is approximated as\(^{18,116}\)

\[ C^{uv}_{\alpha\beta} \approx F_{\alpha X} \langle \varphi_{X} \varphi_{\beta} | F_{12} | \varphi_{u} \varphi_{v} \rangle + \langle \varphi_{\alpha} \varphi_{X} | F_{12} | \varphi_{u} \varphi_{v} \rangle F_{X \beta}. \]  \hfill (86)

The second term in equation (84) is the external exchange operator. It includes all contractions of the doubles amplitudes that involve three or four virtual orbitals and is of the form

\[ [K(D^{uv})]_{\alpha\beta} = \langle \varphi_{\alpha} \varphi_r | \varphi_{\beta} \varphi_s \rangle D_{rs}^{uv}, \]  \hfill (87)

where \(D_{rs}^{uv}\) are composite amplitude matrices, for which \(D_{\gamma\eta}^{uv} = T_{\gamma\eta}^{uv} + c_{\gamma\eta}^{uv}\), \(D_{\eta\gamma}^{uv} = \delta_{\eta\gamma} T_{\gamma\gamma}^{uv}\) and \(D_{\gamma\gamma}^{uv} = 0\). The greek letters \(\gamma\) and \(\eta\) refer to virtual orbitals.

The corrections to \(K(D^{uv})\) arising from the added explicitly correlated terms constitute the third term. With the help of equation (74) the elements of this matrix can be written as

\[ [K(T^{uv})]_{\alpha\beta} = V_{\alpha\beta}^{mn} T_{mn}^{uv}, \]  \hfill (88)

where

\[ V_{\alpha\beta}^{mn} = \langle \varphi_{\alpha} \varphi_{\beta} | r_{12}^{-1} \hat{Q}_{12} F_{12} | \varphi_{m} \varphi_{n} \rangle. \]  \hfill (89)

Using the resolution of identity, this three electron integral can be written as\(^{17,19}\)

\[ V_{\alpha\beta}^{mn} = \langle \varphi_{\alpha} \varphi_{\beta} | \frac{F_{12}}{r_{12}} | \varphi_{m} \varphi_{n} \rangle \\
- \langle \varphi_{\alpha} \varphi_r | \varphi_{\beta} \varphi_s | F_{12} | \varphi_m \varphi_n \rangle - \langle \varphi_{\alpha} \varphi_u | \varphi_{\beta} \varphi_x | F_{12} | \varphi_m \varphi_n \rangle \\
- \langle \varphi_{\alpha} \varphi_x | \varphi_{\beta} \varphi_u | F_{12} | \varphi_m \varphi_n \rangle. \]  \hfill (90)
After the $R_{MP2-F12}^{uv}$ terms, the $K(T^{uv})$ have the largest impact on the amplitudes because these are the only terms that have contractions of doubles amplitudes with integrals over three or four external orbitals.\cite{18}

The final terms in equation (84) consists of terms $\alpha_{uv,kl}$ and $G_{uv}^{\alpha \beta}$. These are intermediates that depend on amplitudes and integrals with at most two external orbitals. The external-external block matrices have the elements $[D^{uv}]_{\alpha \beta} = D_{\alpha \beta}^{uv}$.

The bottleneck of the CCSD(T)-F12 calculation is the coupling of the new explicitly correlated terms with the original CCSD amplitudes. A direct application of the presented CCSD(T)-F12 method increases the computational effort compared with the standard CCSD(T) calculation by an order of magnitude. To avoid this increase, the CCSD(T)-F12\cite{19} introduces several approximations, the most important of which is that it ignores all contributions of the explicitly correlated configurations to $R_{CCSD-F12}^{uv}$ except for $K(T^{uv})$ and the coupling matrices $C_{mn}^{\alpha \beta}$ in equation (85). Furthermore the problematic last two terms in equation (90) involving integrals over two virtual orbitals and one complementary auxiliary orbital can be shown to be small by density fitting procedures\cite{119} and numerical estimation,\cite{18} and are therefore omitted. In this case, the $V_{mn}^{uv}$ terms can be written simply as

$$V_{mn}^{rs} = W_{mn}^{rs} - K(F_{mn}^{rs})$$

where $x$ and $y$ are some orbitals representable in the atomic orbital basis. Because the last term in equation (91) is of the same form as the external exchange operators, the total residual of the CCSD(T)-F12a method simplifies to

$$R_{CCSD-F12}^{uv} = R_{MP2}^{uv} + C_{uv}^{\alpha \beta} + W_{uv} + K(D_{uv}^{\alpha \beta} - F_{uv}^{\alpha \beta})$$

where $F_{uv}^{\alpha \beta} = C_{mn}^{uv}T_{mn}^{uv}$ and the terms $W_{uv}^{\alpha \beta}$ and $C_{uv}^{\alpha \beta}$ are defined analogously.

Once the amplitudes are known, the energies can be calculated. After this, the perturbative triples are handled exactly as in the CCSD(T) method and their energy is added to the F12 energy.\cite{18}

Resulting from all these rather technical approximations, the CCSD(T)-F12a scales formally like the CCSD(T)-method, but gives much more accurate results. With minimal additional computational cost, calculations performed at aug-cc-pVDZ level are often comparable to calculations performed at CCSD(T)/aug-cc-pVQZ level, while an aug-cc-pVTZ basis set gives results similar to CCSD(T)/aug-cc-pV5Z calculations.\cite{18,19,120} The method reduces basis set errors of atomization energies, reaction energies, electron affinities, ionization potentials, equilibrium structures and vibrational frequencies all by an order of magnitude.\cite{18} It has been demonstrated that similar advantages are also found in equilibrium geometries and anharmonic vibrational frequencies for large molecules.\cite{121} For these small basis sets the good performance is probably the result of a cancellation of errors regarding overshooting the F12 correlation energies and undershooting the noncorrected triples contribution.\cite{18} However, as this cancellation is systematic the results are reliable to a high accuracy.
2.2.6 Basis Functions

In quantum chemical calculations, the choice of a set of basis functions is as integral a part of the computational procedure as is the choice of the method used. A good choice of basis functions can reduce computational costs while maintaining the required accuracy, and a bad one can lead to unreliable results.

The two most common types of basis function are the Slater type orbitals (STOs) and the Gaussian type orbitals (GTOs). STOs are made of an exponentially decreasing radial dependence multiplied the angular part of the analytic hydrogen–like orbitals $Y_{l,m}(\theta,\phi)$:

$$\xi_{nlm}(r,\theta,\phi) = N r^{n-1} e^{-\zeta r} Y_{l,m}(\theta,\phi), \quad (93)$$

where $\zeta$ is a parameter that is usually obtained by fitting the STOs to numerically computed atomic orbitals and $N$ is a normalization constant. The variables $(r,\theta,\phi)$ are given in spherical coordinates and for the quantum numbers $l$ and $m$, $l \in \mathbb{N}$ and $m \in \mathbb{Z}$. As can be seen, the Slater type orbitals have no radial nodes except for the one located at origin. It follows a radial nodal structure can only be constructed by taking a linear combination of STOs. The application of STOs is severely limited by the fact that the numerous three- and four-center two electron integrals required by virtually all correlation methods cannot be analytically calculated with them. Due to this difficulty the Slater type orbitals can only be used in the smallest of systems where high accuracies are required or the problematic integrals are ignored altogether.

The Gaussian type orbitals have the general form

$$\xi_{ijk}(x,y,z) = (x-x_c)^i (y-y_c)^j (z-z_c)^k e^{-|\zeta(r-r_c)|^2}, \quad (94)$$

where the point $r_c = (x_c,y_c,z_c)$ defines the center of the Gaussian function and the parameter $\zeta$ determines how fast the GTO declines. Like in the STOs, $\zeta$ is often determined by variational calculations with atoms. Usually the centers of the GTOs used correspond to the nuclei in the system. The natural numbers $i$, $j$ and $k$ define the number of nodes the Gaussian function has. Their sum also classifies the GTOs type as s, p, d etc. It follows from definition (94) that the product of two Gaussian functions gives a Gaussian function, which greatly improves computational efficiency for the numerous two electron integrals encountered in the different correlation methods.

The downside with the GTOs is that owing to the $r^2$ dependence in the exponential term, they often decay too fast compared with the actual wave functions, and also have a zero slope at the center of the Gaussian. Thus, these functions often give a poorer representation of the orbitals, and so a larger basis set is required to achieve comparable accuracy. The increased efficiency in integral calculations, however, vastly outshines these negative aspects of the GTOs. The GTOs are by far the most common type of basis function employed in quantum chemistry.

Variational calculations used to determine the set of parameters $\zeta$ for a set of GTO-basis functions usually give biased results in the sense that the resulting functions are mainly optimized around the chemically uninteresting core electrons of the system. This phenomenon is due to the large contribution
of the core electrons to the system energy.\textsuperscript{65} To circumvent this problem, a set of primitive Gaussian functions of equation (94) centered at the same point are often grouped together into Gaussian contractions (CGTO) of the form
\[ \xi_{g}^{p} = \sum_{l} d_{lp} \xi_{l}^{g}, \]  
where the coefficients \( d_{lp} \) are predetermined and so remain constant during the actual variational calculation.

The application of CGTOs makes the computations a lot cheaper, as the generally unchanging core orbitals can be represented by a single contraction and need not be evaluated again at every cycle of iteration. Most of the modern high-accuracy calculations employ basis sets where the contraction is general, meaning that most of the primitives enter any given contracted Gaussian function but with different contraction coefficients for different CGTOs.\textsuperscript{126} After the contraction, a linear combination of the atom centered CGTOs is used to represent the spatial orbitals of the system as in equation (34) and the coefficients are determined by the Hartree Fock calculation according to section 2.2.2.

The number of basis functions chosen naturally depends on the required accuracy and the complexity of the system. In the minimum basis set, each of the atomic orbitals in the elementary valence theory are represented by one function only. When correlation effects are concerned, this type of basis set almost always gives insufficient results. The Double Zeta (DZ) -type basis sets double the number of basis functions, but in many systems it is necessary to go further to Triple, Quadruple or even Quintuple Zeta basis sets with three, four and five times the basis functions of the minimum basis. Because the computational effort with many of the correlation methods grows rapidly with the basis set size, it is common to use split valence basis sets where the number of basis functions is increased only for the valence electrons. This approach is justified by the fact that the inner shells are almost independent of the chemical environment of the system.

With correlation methods, oftentimes just adding more functions of the same type to a given basis does not significantly improve the computational results. This is because these functions are unable to account for the deformation of the atomic orbitals caused by adjacent atoms and for the angular correlation arising from situations where electrons are on the opposite sides of the nucleus.\textsuperscript{65} This problem can be solved by introducing polarization functions which are functions with a higher value of angular momentum.

It is desirable that the shift from one basis set to a more complete one increases the percentage of correlation energy obtained in some predictable manner. This is achieved in the correlation consistent (cc) basis sets of Dunning,\textsuperscript{32,33} where angular momentum functions that have a similar kind of contribution to the energy are added together as the basis approaches completeness. For example, instead of inserting all the d functions into the basis set at once, the cc basis sets add the polarization functions in the succession 1d, 2d1f, 3d2f1g.\textsuperscript{127}

The application of correlation consistent basis set also makes it theoretically possible to achieve the complete basis limit, i.e., the correlational energy if the basis set used to expand the wave function was infinite. This is done by doing calculations with an increasing amount of basis functions and then
extrapolating the results to the basis set limit. Unfortunately, this method is not feasible in large systems as most of the extrapolation formulae usually have three or more parameters,\textsuperscript{65,128,129} and so a calculation of at least the correlation consistent quadruple zeta basis set is required.

If the system contains hydrogen bonding or strong van der Waals interactions it is necessary to further refine the basis set by adding diffuse functions. Diffuse functions have a small value of the exponential coefficient $\zeta$ in equation (94) and so decay slowly with increasing distance from the Gaussian center. These functions are added to the augmented correlation consistent (aug-cc) variant of the Dunning cc-basis set.\textsuperscript{34} In addition to the correlation consistent basis set described above, other popular basis sets include Ahlrich type basis sets, Pople style basis sets and Dunning and Huzinaga basis sets.\textsuperscript{65}

2.2.7 The Solutions to the Nuclear Hamiltonian

We now turn our attention to the solution of the Nuclear Hamiltonian of equation (30) and assume that the points on the potential energy surface $E_i(y)$ can be calculated as necessary. Because our ultimate goal is the accurate prediction of thermodynamical properties, in this section we shall focus on the calculations of equilibrium geometries and vibrational states. In general, most of the molecular properties can be thought of as the responses of the wave function, energy or an expectation value to a perturbation.

Geometry Optimization

To calculate thermodynamical properties, the equilibrium structures and their energies have to be determined. As the equilibrium geometry corresponds to a global energy minimum of the potential energy surface, the problem is one of locating the minimum energy. Most of the location methods use the property that for a change in the nucleus position $y - y_0$, the change in energy can be written as a Taylor series

$$E(y) = E(y_0) + \left( \frac{\partial E}{\partial y} \right)^T (y - y_0) + \frac{1}{2} (y - y_0)^T \frac{\partial^2 E}{\partial y^2} (y - y_0) + \ldots, \quad (96)$$

where the first derivative is the gradient vector, which points to the direction of the greatest increase in energy, the second derivative is a matrix of the harmonic force constant and higher derivatives are anharmonicity corrections to the vibrational frequencies. The superscript $T$ denotes transpose vector.

Different algorithms for finding the minima can be categorized with respect to equation (96). The crudest methods which make use only of the energies are slowest to converge, but they are useful when the calculation of derivatives is for some reason impossible. The simplex algorithm is a well-known example of such methods. Methods that use the first derivative in addition to the energies are almost an order of magnitude more efficient. Popular examples include the steepest descents method and the conjugate gradients minimization. The most efficient and accurate methods make use of the energy and both the first and second derivatives. The simplest second-derivative method is the Newton-Rhapson method.\textsuperscript{72,125}
Harmonic Wavenumbers

Let us now look at the oscillatory motion of the nuclei around the equilibrium structure. For a deviation from equilibrium with the nuclear coordinates $y_e$, the first derivative of equation (96) disappears because of the stationary nature of $y_e$. By choosing the zero of potential energy to be $E(y_e)$ and ignoring all terms of third and higher order in equation (96), substitution into the nuclear Schrödinger equation (30) gives

$$
\left[ -3N_n \sum_{j=1}^{3N_n} \left( \frac{\hbar^2}{2m_{a,j}} \partial^2 \right) + \frac{1}{2} (y - y_e)^T \frac{\partial^2 E}{\partial y^2} (y - y_e) \right] \psi_n = E_n \psi_n. \quad (97)
$$

This equation can then transformed into mass-dependent coordinates $z_j = \sqrt{m_{a,j}} (y_j - y_{e,j}) / \hbar$:

$$
\left[ -3N_n \sum_{j=1}^{3N_n} \left( \frac{1}{2} \partial^2 \right) + \frac{1}{2} (z)^T (F \cdot G) (z) \right] \psi_n = E_n \psi_n \quad (98)
$$

with the $G$ matrix defined by $G_{jk} = 1 / \sqrt{m_{a,j} m_{a,k}}$ and $F$ is a $3N_n \times 3N_n$ matrix of the force constants ($\partial^2 E/\partial z^2$). Finally, a unitary transformation $U$ is used to diagonalize the $F \cdot G$ matrix. With a final coordinate change $q = Uz$ the Schrödinger equation is transformed into a set of $3N_n$ one-dimensional harmonic oscillator equations:

$$
\left[ -3N_n \sum_{j=1}^{3N_n} \left( \frac{1}{2} \partial^2 \right) + \frac{1}{2} \varepsilon_j q_j^2 \right] = E_n \psi_n. \quad (99)
$$

From the eigenvalues of the unitary transformation $\varepsilon_j$, the harmonic frequencies are obtained by $\nu_j = \sqrt{\varepsilon_j / 2 \pi}$ and the eigenvectors $q$ are the mass-weighted normal coordinates. Once the frequencies are known, the wavenumbers can be obtained from $\tilde{\nu}_j = \nu_j / c$ where $c$ is the speed of light in vacuum.

Theoretically, for a nonlinear molecule, six of the $3N_n$ eigenvalues should be zero (for a linear molecule, the number is five). However, because the equilibrium geometry is only approximate and the gradient is not in truth zero, the six degrees of freedom corresponding to the translations and rotations of the molecule have to be projected out of the calculations. The physical interpretation of the normal modes is such that these correspond to somehow isolated motions of groups of atoms in the sense that each normal mode can be excited without exciting any of the other modes.

With an accurate PES, the harmonic frequencies tend to overestimate the experimental ones by around 5% due the lack of higher-order terms in equation (97). Furthermore, substantial errors are caused by the missing correlation contributions due to the truncation of both the one-electron and $N_e$-electron basis sets.

Anharmonic Vibrations

The harmonic approximation works well in systems where the true potential curve is well approximated by a second order polynomial. This means
that in systems with a single deep minima, the harmonic approximation can be used for the few lowest frequencies with some accuracy. Problems are bound to arise when the vibrational quantum number \( v \) increases, because the equally spaced harmonic ladder fails to describe the ever increasing density of vibrational states as the dissociation limit becomes closer. Also in systems with multiple potential energy minima the splitting of states caused by tunneling effects causes the harmonic approximation to fail completely. In such cases, the higher order derivatives in equation (96) cannot be ignored, and as a result the separation into different normal coordinates is not as useful as before.

As the size of the system increases, the dimensionality of the PES increases by three with each added atom. Unfortunately, the complete anharmonic treatment requires calculations of vast areas of this PES, and so it is already impossible for systems such as the sulfuric acid and sulfuric acid monohydrate complex which have twelve and twenty four vibrational degrees of freedom, respectively. For small systems such as a single water molecule with its three vibrational degrees of freedom, the complete anharmonic treatment of the potential energy surface is doable by using the same general methods as in the solution of the electron correlation problem.

With large molecules, it is possible to work around the difficulties in the anharmonic treatment by dividing the large systems into smaller uncoupled sub-systems (domains) that can then be treated with higher accuracy, a bit like in the local approximations described earlier. From now on this will be known as the anharmonic domain approximation. Two anharmonic domains were used in this work. The first one consisted of those large amplitude modes that could result in a change of configuration and the second consisted of the high frequency vibrations that have the largest effect on the zero-point energies. These high frequency vibrations correspond to the two water stretches in the sulfuric acid monohydrate. Because these are closely coupled to the water bending motion, it was also included in the anharmonic domain. The rest of the vibrational degrees of freedom were treated harmonically.

In the high frequency case the three-dimensional potential energy surface for the water in the monohydrate complex was calculated by keeping the coordinates of the sulfuric acid constant and varying the internal coordinates of the water. Following the approach taken by Kauppi and Halonen,\textsuperscript{130} after the PES had been found the vibrational problem of the water molecule was solved variationally.

For the variational calculation the nuclear Hamiltonian was expressed in terms of the curvilinear internal coordinates. With the omission of angular momentum components in the Hamiltonian, it can be written in the form\textsuperscript{130,131}

\[
\hat{H} = \hat{T} + \hat{V} = -\frac{\hbar^2}{2} \sum_{ij} \left[ \left( \frac{\partial g^{(q_i,q_j)}}{\partial q_i} \right) \frac{\partial}{\partial q_j} + g^{(q_i,q_j)} \frac{\partial^2}{\partial q_i \partial q_j} \right] + \hat{V}'(q) + \hat{V}(q), \quad (100)
\]

where \( q \) are the curvilinear internal coordinates. In equation (100), \( g^{(q_i,q_j)} \) quantities are the elements of the mass-weighted reciprocal matrix and have
the form \[ g(q_i) = \sum_{\alpha} \frac{1}{m_{\alpha}} (\nabla_\alpha q_i) \cdot (\nabla_\alpha q_j), \] (101)

where \( m_{\alpha} \) are the masses of the nuclei. The Jacobian of the coordinate transformation is denoted \( J \) and it appears in \( V'(q) \) which is a small pseudopotential term that is independent of the momentum operators\(^{132,133} \) and is of quantum mechanical origin. The quantity \( V'(q) \) is the potential energy operator given in the curvilinear internal coordinates. The volume element for integration is such that the weight function is one.

The two major advantages associated with the use of curvilinear coordinates are that these give a more accurate representation of the potential energy surface and make the potential energy surface parameters independent of the isotopes within the Born-Oppenheimer approximation.\(^{132} \) The one major drawback with the curvilinear coordinates is that kinetic energy operator becomes more complicated in curvilinear internal coordinates as can be seen from equation (100).

As in the original work by Halonen and Carrington,\(^{132} \) the potential energy operator was expressed in terms of the Morse variables \( z_b \) and \( z_f \) for the stretches and the curvilinear internal displacement coordinate \( \theta \) for the bend. The Morse variable \( z_b \) is defined by the equation \( z_b = 1 - e^{-q_b r_b} \), where \( r_b = R_b - R_{b,e} \), and \( a_b \) is the Morse parameter.\(^{134} \) \( R_b \) is the instantaneous bond length of the hydrogen bonded hydrogen in the water molecule and \( R_{b,e} \) is the equilibrium bond length. For the bond between the free hydrogen and the oxygen, \( z_f \) is defined analogously. The displacement coordinate is defined by \( \theta = \phi - \phi_e \), where \( \phi \) is the instantaneous valence angle and \( \phi_e \) is its equilibrium value.

For systems such as the isolated water molecule, it follows from symmetry that \( R_{b,e} = R_{t,e} \) and \( a_b = a_t \) for the Morse parameters. However, due to the asymmetry of the sulfuric acid environment, the hydrogens on H\( O \) in the sulfuric acid monohydrate have different equilibrium bond lengths and Morse parameters. To account for this, a more general form of the potential energy in Ref.\(^{130} \) was derived for this study by comparing the different derivatives of the Taylor expanded potential energy function in the \( r_b, r_t \) and \( \theta \) coordinates to the derivatives of the potential energy operator

\[
\hat{V}(z_b, z_f, \theta) = F_{z_b, z_b} z_b^2 + F_{z_f, z_f} z_f^2 + F_{\theta \theta} \theta^2 + F_{z_b, z_f} z_b z_f + F_{z_f, \theta} z_f \theta + \ldots \]

(102)
at the point of equilibrium where \((r_b, r_t, \theta) = (z_b, z_f, \theta) = 0\). After a straightforward but rather lengthy calculation, the following form could be obtained for the potential energy operator

\[
V(z, z_f, \theta) = \sum_{k=2}^8 T_k^{(b)} z_b^k + \sum_{k=2}^8 T_k^{(f)} z_f^k + \frac{f_{r_b r_t}}{a_b a_t} z_b z_f + \frac{1}{2} \left( \frac{f_{r_b r_f}}{a_b a_t} \right) z_b^2 z_f + \frac{1}{2} \left( \frac{f_{r_t r_f}}{a_b a_t} \right) z_f^2 z_b + \frac{f_{\theta \theta}}{2 a_b a_t} z_b^2 \theta^2 + \frac{f_{r_t \theta}}{2 a_f} z_f^2 \theta^2 + \ldots
\]

\[
+ \sum_{k=2}^4 \frac{f_{\theta \theta}}{k!} \theta^k + \frac{f_{r_b \theta}}{a_b} z_b \theta + \frac{f_{r_t \theta}}{a_f} z_f \theta + \frac{f_{r_b \theta \theta}}{2 a_b} z_b \theta^2 + \frac{f_{r_t \theta \theta}}{2 a_f} z_f \theta^2
\]

30
The potential energy term from the Hamiltonian used in the calculation can be found elsewhere. Given by

\[ H = \frac{1}{2a_b^2} (f_{r, \tau} + a_b f_{\tau}) \dot{z}_b^2 + \frac{1}{2a_f^2} (f_{r, \tau} + a_f f_{\tau}) \dot{z}_f^2 + \frac{1}{4a_b^2} (f_{r, \theta, \tau} + a_b f_{\theta, \tau}) \dot{z}_b^2 + \frac{1}{4a_f^2} (f_{r, \theta, \tau} + a_f f_{\theta, \tau}) \dot{z}_f^2. \] (103)

The above equation is generally of the same form as the potential energy operator used in Ref.\textsuperscript{130} As can be seen, a polynomial up to eighth order was used in the Morse variables while only a fourth order polynomial was needed for \( \theta \). Coupling terms between the stretches and the bend were included up to fourth order and up to third order between the two stretches.

From equation (100) the eigenvalues were obtained variationally. A basis set of harmonic oscillator wave functions \( \xi^h(\theta) \) of the type

\[ \xi^h(\theta) = \left( \frac{\alpha}{2^v v! \pi^2} \right)^{1/2} H_v(\alpha \theta) e^{-\alpha^2 \theta^2/2} \] (104)

was used for the bend. In equation (104) \( \alpha \) is a parameter that depends on the form of the Hamiltonian\textsuperscript{28} and \( H_v(\alpha \theta) \) is a Hermite polynomial.\textsuperscript{73} Morse oscillator eigenfunctions of the form\textsuperscript{134}

\[ \xi^m(x) = n! \left[ \frac{a(2\lambda - 2n - 1)}{\Gamma(n + 1) \Gamma(2\lambda - n)} \right]^{1/2} x^{\lambda - n - 1/2} e^{-1/2 x^2} L_n^{2\lambda - 2n - 1}(x) \] (105)

were used for the stretches. In equation (105), \( L \) is the Laguerre polynomial,\textsuperscript{73} \( \Gamma \) is the gamma function,\textsuperscript{135} and \( x = 2\lambda e^{-\alpha(R - R_s)} \) where \( \alpha \) is the Morse parameter. The parameter \( \lambda \) is related to the depth of the potential well \( D_b \) via \( \lambda = \sqrt{2mD_b/\hbar} \). The total wave function was expressed as a linear combination of these functions

\[ |\psi\rangle = \sum_{n_b} \sum_{n_f} \sum_v c_{n_b n_f v} \left( \xi^m_{n_b}(x_b) \right) \left( \xi^m_{n_f}(x_f) \right) \left( \xi^h(\theta) \right), \] (106)

where the coefficients \( c_{n_b n_f v} \in \mathbb{R} \). The analytical matrix elements resulting from the Hamiltonian used in the calculation can be found elsewhere.\textsuperscript{130,132}

Two of the large amplitude motions of the sulfuric acid monohydrate were considered for anharmonic treatment. These were chosen on the basis that these could result in a transformation from one conformer to another. The motions were the torsional rotation of the non-hydrogen bonded hydrogen on the sulfuric acid and the wagging motion of the free hydrogen on the water molecule. See section 3.1 and figure 2 for details.

The problem was solved variationally with the Hamiltonian for the system given by

\[ \hat{H} = -\frac{\hbar^2}{2I_\tau} \frac{\partial^2}{\partial \tau^2} - \frac{\hbar^2}{2I_\theta} \frac{\partial^2}{\partial \theta^2} + V(\tau, \theta), \] (107)

where \( \tau \) is the torsional angle and \( \theta \) is the wagging angle. The moments of inertia \( I_\tau \) and \( I_\theta \) are defined by

\[ I_i = m_H r_{H,i}^2 \sin \theta_{H,i}, \] (108)

where \( m_H \) is the mass of hydrogen, \( r_{H,i} \) is the bond length of the hydrogen, \( \theta_{H,i} \) is the angle between the bond and the rotational axis and \( i = \tau \) or \( \theta \).

The potential energy term \( V(\tau, \theta) \) is of the form

\[ V(\tau, \theta) = A_0 + \sum_{m=1}^{6} (A_m \cos m \tau + B_m \sin m \tau) + \sum_{k=2}^{8} C_k \theta^k. \]
\[ f \tau \varphi (\tau - \tau_e) (\vartheta - \vartheta_e) + f \tau \vartheta (\tau - \tau_e)^2 (\vartheta - \vartheta_e) + f \tau \varphi (\tau - \tau_e) (\vartheta - \vartheta_e)^2 \]
\[ + f \varphi \vartheta (\tau - \tau_e)^2 (\vartheta - \vartheta_e)^2 + f \tau \varphi \vartheta (\tau - \tau_e) (\vartheta - \vartheta_e)^3 \]
\[ + f \tau \varphi \vartheta (\tau - \tau_e)^3 (\vartheta - \vartheta_e), \quad (109) \]

where \( \tau_e \) is the equilibrium value for the torsional angle, and \( \vartheta_e \) is the equilibrium value for the wagging angle. As can be seen, a Fourier series up to the sixth order was used to describe the one-dimensional part in the torsion angle \( \tau \), while a simple eighth order polynomial was used for \( \vartheta \). Coupling terms were included up to fourth order. The wavefunction \( \psi \) of the system was approximated by a harmonic oscillator basis set \( \xi^h(\vartheta) \) for wagging angle and a linear rotor basis set of the type

\[ \xi_k^r(\tau) = \frac{1}{\sqrt{2\pi}} e^{ik\tau} \quad (110) \]

for the torsional motion. The quantum number \( k \) has values \( k \in \mathbb{Z} \). As before, the wave function was represented as a linear combination of the product of these functions

\[ |\psi\rangle = \sum_k \sum_v c_{kv} |\xi_k^r(\tau)\rangle |\xi_v^h(\vartheta)\rangle, \quad (111) \]

where the coefficients \( c_{kv} \in \mathbb{C} \).

Insertion of the wave function (111) into the Schrödinger equation with the Hamiltonian defined by equation (107) followed by multiplication from the right with \( \langle \xi_k^r(\tau) | \langle \xi_v^h(\vartheta) | \), leads to a matrix equation of the form

\[ \sum_k \sum_v c_{kv} \langle \xi_v^h(\vartheta) | \langle \xi_k^r(\tau) | \hat{H} | \xi_k^r(\tau) \rangle | \xi_v^h(\vartheta) \rangle = c_{k'v'} E, \quad (112) \]

from which the energies can be obtained by diagonalizing the matrix.

### 2.3 Rudimentary Statistical Thermodynamics

Statistical thermodynamics provides the link between microscopic quantities quantum mechanics is concerned with, and macroscopic quantities covered by thermodynamics. This section deals with the basics of statistical thermodynamics, introducing the formulae for enthalpies and equilibrium constants which will be applied to the sulfuric acid–water monomer system in sections 3.4 and 3.5.

#### 2.3.1 Ensembles and Partition Functions

The macroscopic state of a large system, with the number of particles ranging in moles, can be defined by specifying a limited number of the thermodynamic quantities of the system such as \( n, T \) and \( p \). On the other hand, the identification of the microscopic state of such a system would require the specification of all quantum numbers of the system, and is utterly unfeasible. Thus, one of the fundamental problems in statistical thermodynamics is that to calculate anything from the microscopic quantities of quantum mechanics, we would have to specify the state of the system.
Another fundamental problem arises because the measuring process of thermodynamic quantities takes a finite amount of time, the results are in essence time averages of the determined properties. Theoretical calculation of such a time average often requires that both the momenta and the locations of the particles be determined after each time step – again an overwhelming task for large systems.

To deal with the problem of time averages, statistical mechanics introduces the concept of an ensemble, which is an arbitrarily large collection of systems that are all in the same macrostate but with different microstates. Because of the arbitrary nature of the ensemble, it can be constructed such that it consists of all possible microstates. The first postulate of statistical thermodynamics is that the ensemble average of such an ensemble corresponds to the time average obtained as a result from a thermodynamical measurement.

The problem in the specification of the system state is harder to resolve. To begin with, a second postulate of equal \textit{a priori} probabilities is necessary. The postulate states that because we know nothing of our system beforehand, we assign an equal probability for each of the different microstates of the system, i.e., we assume that each of the systems appear equally many times in the ensemble. In this context, the system is thought to have constant $E$, $V$ and $N$ and degeneracy $\Omega(N,V,E)$. With these two postulates, the value of a macroscopic thermodynamic variable $X$ for any ensemble is simply

$$\bar{X} = \sum_j X_j P_j,$$

where $P_j$ is the average probability of finding a system in a certain energy state where the value of $X$ is $X_j$.

Let us look at the problem of obtaining $P_j$ in an ensemble where $T$, $V$ and $N$ in each system are kept constant. This is called a canonical ensemble and it can be visualized as consisting of an infinite series of labeled closed systems immersed in a constant temperature bath. Naturally, the largest contribution to the average probability comes from the term that can be achieved in the greatest number of ways, i.e., the average probability will be most affected by the term for which the weight $W(a)$ is maximum

$$W(a) = \frac{\mathcal{A}!}{\prod_k a_k!},$$

where $a_k$ is the number of systems in the ensemble with the energy $E_k$ and $\mathcal{A}$ is the total number of systems in the ensemble. It turns out that as the size of the ensemble increases, the term that can be achieved in the greatest number of ways so overwhelms the others that the average probability can be replaced by the one corresponding to the most likely state of the entire ensemble.

For different ensembles with different conditions on the systems they contain, the conditions under which equation (114) is maximized change. In addition to the canonical ensemble both microcanonical and grand canonical ensembles are commonly employed. In the microcanonical ensemble, the constants are $E$, $V$ and $N$ and in the grand canonical ensemble they are $T$, $V$ and the chemical potential $\mu$.

Regardless of the ensemble used, the maximization leads to the concept of a partition function. The partition functions for the three ensembles are
represented below \(^{27}\)

\[
M(N,V,E) = \Omega(N,V,E) \tag{115}
\]

\[
Q(N,V,T) = \sum_k \Omega(N,V,E_k)e^{-\frac{E_k(N,V)}{kT}} = \sum_j e^{-\frac{E_j(N,V)}{kT}} \tag{116}
\]

\[
\Xi(N,V,\mu) = \sum_N Q(N,V,E)e^{-\frac{\mu N}{kT}}, \tag{117}
\]

where \(\Omega(N,V,E)\) is the degeneracy of the \(E\)th energy level. In equation (116), the sum in the middle term is over the different energy levels and the sum on the right is over all the energy states of the system.

Because all of the fundamental thermodynamic properties can be relatively easily represented in terms of the partition function, it is of central importance in statistical mechanics. For example from the canonical partition function, the Helmholtz energy can be simply obtained from

\[
A = -kT \ln Q(N,V,E). \tag{118}
\]

The connection between equations (115) through (117) and the different ensembles is mathematically rather simple. Like the Helmholtz energy is a Legendre transformation of the internal energy with respect to the entropy of the system, so is the canonical ensemble a Legendre transformation of the microcanonical one. The grand canonical ensemble is nothing more than a Legendre transform of the canonical ensemble.

It should be noted that the direct evaluation of the partition functions (116) and (117) still requires the complete knowledge of the system’s energy levels. To make use of these equations in practice, a number of simplifying assumption have to be made.

In low enough densities and high temperatures, the gas molecules can be treated as ideal. This makes it possible to separate the Hamiltonian for the system into a sum of single molecule Hamiltonians and the energy of the state \(E_j\) can be represented as a sum of the molecular energy levels \(\epsilon\)

\[
Q(N,V,T) = \sum_{a,b,c,d,\ldots} e^{-\frac{\epsilon_a + \epsilon_b + \epsilon_c + \ldots}{kT}}. \tag{119}
\]

If the particles are distinguishable this sum is easy to evaluate as the summing can be done separately for each \(i = a,b,\ldots\). However, with indistinguishable particles it becomes difficult when two or more particles occupy the same molecular state. Luckily for gas systems at roughly room temperature, the number of permissible translational states is usually so large that their number vastly outnumbers the number of molecules in the system.\(^{27}\) As a result, it becomes extremely improbable to find a multiply occupied state and the canonical partition function becomes

\[
Q(N,V,T) = q^N \tag{120}
\]

where \(q\) is the molecular partition function

\[
q = \sum_i e^{-\frac{\epsilon_i}{kT}}. \tag{121}
\]
With this, the problem of calculating the energy states for the whole system has been transformed into a calculation of all the energy states of a single molecule, which is a much simpler problem. Equation (121) ignores all the additional requirements the wavefunction may have, such as the antisymmetry requirements for fermions. Accounting for these effects leads to a different form of the canonical partition function, but it can be shown that in the Boltzmann limit of high temperature and low density, all these partition functions reduce to equation (121).

2.3.2 The Evaluation of Partition Functions

It seems obvious that to evaluate the molecular partition function $q$, some additional approximations are necessary. We make the assumption that the Hamiltonian can be further separated according to the different degrees of freedom, and the electronic and nuclear contributions so that

$$\hat{H} = \hat{H}_t + \hat{H}_r + \hat{H}_v + \hat{H}_e + \hat{H}_n.$$  \hfill (122)

The subindex $t$ refers to translational motion, $r$ refers to rotational motion, $v$ refers to vibrational motions, $e$ refers to the electronic hamiltonian and finally $n$ to the rest of the nuclear part of the total hamiltonian. This means that the molecular energy can be represented as a sum of the different terms $\epsilon_t$ through $\epsilon_n$ and the molecular partition function becomes a product of the different terms

$$q = q_t q_r q_v q_e q_n.$$  \hfill (123)

We now justify this partition as well as represent formulas for the different partition functions.

Because the overall motion of the molecule can be separated into the motion of the center of mass of the molecule through space, and the relative motion of the nuclei with respect to the center of mass, it is clear that the translational motion can be treated separately from all the other contributions to the energy. The form of $q_t$ can then be obtained by applying the energies from the particle in a box model to the definition of the molecular partition function (121) with the obvious changes resulting from equation (122), and is simply

$$q_t = \left(\frac{2\pi m kT}{\hbar^2}\right)^{3/2} V,$$  \hfill (124)

where the zero point of the translational energy has been chosen to be that of the lowest translational level and $m$ is the total mass of the molecule.

The electronic and nuclear effects of the total Hamiltonian of the molecule can be separated away from the rest of the degrees of freedom and can, to a good approximation, be treated as separate from one another.\textsuperscript{27} In both of these cases the partition functions can be treated simply according to definition (121). For the electronic partition function, we choose the zero of energy as the energy of the ground state constituent atoms at infinite separation from each other, and for the nuclear partition function it is chosen as the ground state nuclear energy. With this choices, the ground state electronic energy is equal to the depth of the potential energy well $-D_e$ and the electronic partition function is of the form

$$q_e = e^{\frac{-D_e}{kT}} \sum_j \omega_j e^{-\frac{\epsilon_j}{kT}},$$  \hfill (125)

35
where $\epsilon_j$’s are measured from the ground state of the molecule, the sum is over the electronic levels and $\omega_j$ is the degeneracy of the level $j$.

A simplification is obtained by noting that due to the large separations of the nuclear states, exceedingly high temperatures of the order $10^{10}$ K are required for the excited states to be significantly populated.\(^{27}\) This means that the nuclear contribution to the partition function consists only of the degeneracy of the ground state. Furthermore, since the nuclear states seldom change in chemical reactions, it becomes permissible to ignore the nuclear contribution altogether.

The temperatures required for the lowest electronically excited states to be significantly populated are much lower than those for the nuclear states, but still often high enough so that $q_e$ can be well approximated simply by the first term in equation (125). For example in the case of diatomic molecules, the introduction of these terms to equation (125) typically becomes important for temperatures over 5000 K.\(^{137}\)

The separation of the Hamiltonian to the rotational and vibrational degrees is a more complicated issue because of the well known interconnectedness of the two motions.\(^{138}\) Equation (122) is therefore only the starting point of accurate calculations, but it turns out that for most problems the corrections can be made by adding some small terms to the equation that account for centrifugal distortion effects, anharmonic effects and others.\(^{27}\)

A simple form of the rotational partition function can be derived by first making the rigid rotor assumption, in which the movement of the molecule around principal axes of inertia has no effect on its shape, and by assuming that the density of the rotational states is so large that the sum in equation (121) can be approximated by an integral. This can be done when the ratio $\frac{\epsilon_i}{kT}$ is small, which is satisfied when $T$ is large enough. With these approximations the rotational partition function becomes

$$q_r = \frac{1}{\sigma} \left( \frac{kT}{\hbar c} \right)^\frac{3}{2} \left( \frac{\pi}{ABC} \right)^\frac{1}{2}, \quad (126)$$

where $\sigma$ is the symmetry number, which is simply the sum of the different rotational symmetry elements of the molecule. $A$, $B$ and $C$ are called rotational constants and they are defined by the formula

$$A = \frac{h}{8\pi c I_A}, \quad (127)$$

where $I_A$ is the moment of inertia around one of the three principal axes. For rotational degrees of freedom, the energies are generally small and the temperatures required for the approximation to be valid are only some tens of Kelvins. Thus, equation (126) produces accurate values of the partition function for ordinary temperatures.

For vibrational degrees, the separation between the energy levels is so large that in most cases the integral approximation is invalid even at room temperature. As demonstrated before, by applying the harmonic oscillator model and treating the system in normal coordinates, it becomes possible to separate the $3N_n - 6$ vibrational degrees of freedom for a non-linear molecule, or $3N_n - 5$ vibrational degrees of freedom for a linear molecule. Thus, the vibrational partition function becomes a product of $3N_n - 6$ independent
harmonic oscillators

\[ q_{\text{vib}} = \prod_{i=1}^{3N_n-6} q_{\text{vib},i} = \prod_{i=1}^{3N_n-6} \frac{e^{-\frac{h\nu_i}{2kT}}}{1 - e^{-\frac{h\nu_i}{kT}}}, \quad (128) \]

where \( \nu_i \) is the frequency of the \( i \)th vibrational degree of freedom and the zero of energy has been chosen as the bottom of the internuclear potential well. With these choices of the zero of energy in the electronic and vibrational partition functions, the zeroes of energies are equal for all different molecules. This has the effect of making the derivations easier on some occasions.

In systems where the harmonic approximation cannot be used, the vibrational wavefunction has to be evaluated directly from equation (121), which is easy enough provided that the sum is convergent and the vibrational states are readily available.

### 2.3.3 Statistical Mechanics of Equilibrium Systems

From standard thermodynamics we know that for a system in equilibrium, the condition

\[ \sum_i v_i \mu_i = 0 \quad (129) \]

has to hold. Here \( \mu_i \) is the chemical potential of the \( i \)th reaction component and \( v_i \) is its stoichiometric coefficient. On the other hand, the chemical potential can be expressed in terms of the canonical partition function as

\[ \mu_i = -kT \left( \frac{\partial \ln Q}{\partial N_i} \right)_{N_i \neq k,V,T}, \quad (130) \]

where \( N_i \) is the number of \( i \) molecules. If the gases are ideal, the partition function \( Q \) for the whole gas mixture that composes the system can be represented as a product of partition functions for different gases \( Q_k \) which are each given by equation (120)

\[ Q = \prod_i Q_i = \prod_i \frac{q_i^{N_i}}{N_i!}. \quad (131) \]

Insertion of equations (131) and (130) into (129) results in an equation for the equilibrium constant.

\[ K = \prod_k \left( \frac{q_{m,k}^0}{N_A} \right)^{v_k}, \quad (132) \]

where \( N_A \) is Avogadro’s number. In equation (132), we have introduced the concept of standard molar partition function \( q_m^0 \), which is the molecular partition function \( q_m = q/n \) evaluated when \( p = p^0 = 1 \text{ bar} \). The molecular partition function can in principle be evaluated with equations (123), (125), (126) and (128). Note that even though \( q_m^0 \) is directly proportional to \( V \), the ideal gas law states that \( V = V(T) = nRT/p^0 \), and we are left with \( K = K(T) \).

We then derive the temperature dependence of \( K \). In standard thermodynamics, this dependence is described by the van’t Hoff equation

\[ \frac{d \ln K}{dT} = \frac{\Delta H_m^0(T)}{RT^2}. \quad (133) \]
\[ \Delta H_m^\circ \] is the standard molar enthalpy of the reaction. This enthalpy can also be connected to the canonical partition function via

\[ H = kT^2 \left( \frac{\partial \ln Q}{\partial T} \right)_{N,V} + kTV \left( \frac{\partial \ln Q}{\partial V} \right)_{N,T}. \]  
(134)

Now let us consider a system with anharmonic domains. For the vibrational degrees of freedom corresponding to the anharmonic domains, the partition function can be calculated from the direct sum of equation (121) with the zero of potential energy chosen as in the case of the harmonic vibrations. The harmonic vibrational degrees of freedom were calculated from equation (128). By performing the differentiations in equation (134), we obtain

\[ \Delta H_m = \Delta r E_m + 4RT \sum_i v_i + N_A h \sum_j v_j e^{-\frac{h\nu_j}{kT}} \]
\[ + N_A h \sum_d v_d \sum_l e^{-\frac{h\nu_dl}{kT}}, \]  
(135)

where the index \( i \) goes over reactants and products, \( j \) goes over all the harmonic vibrational degrees of freedom of the sulfuric acid monohydrate, sulfuric acid and water molecules, \( d \) goes over the anharmonic domains and \( l \) goes over the different states within the anharmonic domain \( d \). The coefficients \( v_i, v_j \) and \( v_d \) are the reaction coefficients and these are negative in the case of reactants and positive in the case of products. From reaction (18) we can see that for the complex these have a value of one, and for water and sulfuric acid these have the value of minus one.

By setting \( p = p^\circ \), we obtain the standard molar reaction enthalpy from equation (135). Insertion of this into the van’t Hoff equation and integration from a reference temperature of \( T_0 = 298 \) K to the temperature \( T \) yields the following equation for temperature dependence

\[ \frac{K(T)}{K(T_0)} = \exp \left[ -\Delta r E_m \left( \frac{1}{T} - \frac{1}{T_0} \right) + 4 \left( \sum_i v_i \right) \ln \left( \frac{T}{T_0} \right) \right] \]
\[ \cdot \exp \left[ - \sum_j v_j \ln \left( \frac{1 - e^{-\frac{h\nu_j}{kT_0}}}{1 - e^{-\frac{h\nu_j}{kT}}} \right) + \sum_d v_d \ln \left( \frac{\sum_l e^{\frac{h\nu_dl}{kT}}}{\sum_l e^{\frac{h\nu_dl}{kT_0}}} \right) \right]. \]  
(136)

Equation (136) is the statistical thermodynamics expression for the temperature dependence of the stability constant with some of the vibrational degrees of freedom treated by anharmonic domains. The indices have the same meaning as in equation (135).

With the molar enthalpy known, we can also use the Gibbs-Helmholz equation

\[ \left[ \frac{\partial (G/T)}{\partial T} \right]_p = -\frac{H}{T^2} \]  
(137)

to derive the temperature dependence of the molar reaction Gibbs free energy. The statistical thermodynamic equation for the Gibbs free energy of an ideal gas can be obtained directly from the equation (118) and the definition of the Gibbs free energy \( G = A + pV = A + nRT \), and is

\[ G = -nRT \ln \frac{q_m}{N_A}, \]  
(138)
The equation we obtained for the molar reaction Gibbs free energy was of the form

$$
\Delta_r G_m(p,T) = \Delta_r G_m(p,T_0) \left( \frac{T}{T_0} \right) + \Delta_r E_m \left( 1 - \frac{T}{T_0} \right) - 4RT \left( \sum_i v_i \right) \ln \left( \frac{T}{T_0} \right)
$$

$$
+RT \sum_j v_j \ln \left( \frac{1 - e^{-\frac{h\nu_j}{kT}}}{1 - e^{-\frac{h\nu_j}{kT_0}}} \right) - RT \sum_d v_d \ln \left( \frac{\sum_l e^{-\frac{h\nu_{dl}}{kT}}}{\sum_l e^{-\frac{h\nu_{dl}}{kT_0}}} \right),
$$

where the pressure $p$ functions as a parameter that defines $\Delta_r G_m(p,T_0)$ via the translational partition function in equation (138).
3 The Computational Results

3.1 Geometry Optimizations and Energies

The first leg of the journey towards finding the equilibrium constants involves the determination of the equilibrium structures of the molecule. In this study, most of the geometry optimizations were done in three stages. First, a DF-SCS-LMP2 calculation with the Dunning\textsuperscript{32-34} aug-cc-pVDZ basis set was performed from an initial guess based on previous studies and chemical intuition. This guess was then refined with the same method using a higher quality aug-cc-pVTZ basis set. All the DF-SCS-LMP2 geometry optimizations made use of the SCS scaled MP2 energy functional. The localization was done using the Pipek–Mezey\textsuperscript{85} method and local domains were determined automatically using the procedure of Boughton and Pulay.\textsuperscript{139}

The final geometries used for the calculation of rotational constants were obtained from a CCSD(T)-F12a calculation with a VDZ-F12 basis of Peterson et al.\textsuperscript{31} that further improved upon the MP2 calculations. The VDZ-F12 basis is optimized for the use in F12 calculations and has a similar size as the aug-cc-pVDZ basis set. The CCSD-F12a/VDZ-F12 geometries were needed for the harmonic frequency calculations done at this level. From this point onward, we shall denote for short the aug-cc-pVDZ basis by AVDZ and the aug-cc-pVTZ by AVTZ. Because both the local and F12 methods greatly decrease basis set superposition errors\textsuperscript{18,88} the counterpoise correction was not done in this study.

In the CCSD-F12a and CCSD(T)-F12a calculations, AVDZ/MP2FIT auxiliary basis sets of Weigend et al.\textsuperscript{102,140} were used in the density fitting, except for the calculation of the exchange and Fock operators where a VDZ/JKFIT basis set was used. For the resolution of identity parts, the VDZ-F12/optri auxiliary basis set of Yusef and Peterson\textsuperscript{141} was used. The correlation factor was of the form given in equation (76) with $\gamma = 1$. This choice of basis has been found by others to work well for molecules of similar kind.\textsuperscript{120,142} All geometry optimizations, single point calculations and frequency calculations were carried out with the MOLPRO suit of programs.\textsuperscript{143}

3.1.1 Water and Sulfuric Acid Molecules

The geometries for both sulfuric acid and water molecules are given in Appendix I. In both cases, the results are compared to experimental values and to the results of computational studies performed by others. Because of the simplicity of the water system, the DF-SCS-LMP2/AVDZ calculation was omitted as unnecessary.

In Appendix I, the OH bond length in water was denoted simply by $r$ and the HOH angle by $\alpha$. We see that as the correlation treatment becomes more sophisticated, values increasingly close to the experimental\textsuperscript{38,144,145} ones are observed, and in terms of geometry, already a CCSD-F12a method provides an excellent agreement. Curiously, the regular MP2/AVTZ treatment by Kim et al.\textsuperscript{37} seems to surpass the DF-SCS-LMP2/AVTZ method in predicting the equilibrium geometries.

Comparing the CCSD(T)-F12a/VDZ-F12 result with the CCSD(T) calculations of DZ and QZ qualities, one sees that the QZ results are close to the
Figure 1: The two sulfuric acid conformers labeled as g1 and g2 in the order of increasing energy. The bond lengths and angles specified refer to Appendix I.

F12, as already shown by Lane and Kjaergaard\textsuperscript{120} for the water monomer. As the basis-set convergence is not achieved on a CCSD(T)/DZP level, this calculation is inferior to all three done in this work.

It should be noted that, except for the water molecule, because the experimental geometry values correspond to the molecules in the ground vibrational state, these are not exactly the same as the equilibrium values obtained in a geometry optimization. It is possible to determine vibrationally averaged geometries\textsuperscript{146} that account for this effect, but this is beyond the scope of this study.

For sulfuric acid, the two stable conformers found by several other authors\textsuperscript{9,13,147} were detected in this study as well. The structures of the different conformers are given in Figure 1. As can be seen, the difference in these configurations is in the direction of the OH-group, i.e., in the value of the $d_{\text{H7O6S3O2}}$ dihedral angle. Of these two, the g1 geometry that has C\textsubscript{2} symmetry was found to be more stable. It can be seen from the figure that the g2 configuration corresponds to C\textsubscript{s} symmetry.

Looking at the table in Appendix I, one can see from the first four columns the general trend that a shift to a larger basis set leads to a decrease in bond lengths and an increase in bond angles. This trend has been reported by others.\textsuperscript{37,65} It likely results from the fact that the decrease in the total energy of the system resulting from a more complete basis translates to tighter, i.e., more energy containing bonds. This decreases in bond lengths then result in the increases in bond angles. In the case of H\textsubscript{2}SO\textsubscript{4}, the average decrease in bond lengths is around $-0.029\ \text{Å}$ and the increase in bond angle is around 0.3 degrees.

On the other hand, an improvement in the correlation method results in increased bond lengths as more and more of the correlational effect is taken into account and the electrons are pushed away from the bonds. This fact is beautifully reflected in the transition from the CCSD-F12a method to the CCSD(T)-F12a one where the average length increase is 0.006 Å.
Table 1: Energies and relative energies for the geometry optimized structures for water and the different sulfuric acid conformers. \( E \) values are reported in hartree and \( \Delta E \) values are in kJ mol\(^{-1}\). The VDZ-F12 basis has been denoted simply as VDZ.

<table>
<thead>
<tr>
<th></th>
<th>DF-SCS-LMP2/AVDZ</th>
<th>DF-SCS-LMP2/AVTZ</th>
<th>CCSD(T)-F12a/VDZ</th>
<th>CCSD-F12a/VDZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( E )</td>
<td>( \Delta E )</td>
<td>( E )</td>
<td>( \Delta E )</td>
</tr>
<tr>
<td>( \text{H}_2\text{O} )</td>
<td>-76.32770</td>
<td>-76.36319</td>
<td>-76.35640</td>
<td></td>
</tr>
<tr>
<td>( \text{H}_2\text{SO}_4 )</td>
<td>-699.05398</td>
<td>-699.41098</td>
<td>-699.59581</td>
<td>-699.55526</td>
</tr>
<tr>
<td>( \text{g}1 )</td>
<td>-699.05171</td>
<td>5.95</td>
<td>-699.40863</td>
<td>6.15</td>
</tr>
<tr>
<td>( \text{g}2 )</td>
<td>-699.05171</td>
<td>5.95</td>
<td>-699.40863</td>
<td>6.15</td>
</tr>
</tbody>
</table>

Compared with the experimental values,\(^{36}\) again the best performance in the table is seen with the CCSD(T)-F12a method, followed by the CCSD(T) (fc)/V(T+d)Z results of Demaison et al.\(^{147}\) where the valence Triple Zeta basis has an extra tight d function added to the sulfur. The CCSD-F12a is superior compared to DF-SCS-LMP2/AVTZ in predicting bond angles, while the opposite is true for bond lengths. In angle prediction the DF-SCS-LMP2/AVTZ outperforms the standard MP2 calculation with a 6-311++G(2d,2p) Pople style polarized Triple Zeta basis set with diffuse functions performed by Beichert and Schrems,\(^{39}\) but the standard MP2 gives better results for the bond lengths. The performance of the cited\(^{10}\) B3LYP hybrid density functional theory method tends to be on par with the MP2 methods, while the other pure PW91\(^9\) density functional shows the worst performance in this group.

The energies of the two calculations are represented in Table 1. The absolute energies are listed in hartrees and the energies relative to the \( \text{g}1 \) sulfuric acid geometry are given in kJ mol\(^{-1}\). We see from the table that the more sophisticated the correlation treatment becomes, the lower the total energy, just as one would expect to be the case. Because of the rather poor treatment of the chemically uninteresting core electrons, different correlation treatments usually give different absolute values (one hartree equals to around 2625.49 kJ mol\(^{-1}\)), even though the energy differences are similar in size. Due to the simplicity of the sulfuric acid system, the DF-SCS-LMP2/AVTZ method is able to predict the same energy difference of 6.15 kJ mol\(^{-1}\) as the CCSD(T)-F12a/VDZ-F12 treatment.

### 3.1.2 Sulfuric Acid Monohydrate

A total of four different stable geometries were found for the sulfuric acid - water monomer and are given in Figure 2. Although all of these conformers have shown up in past studies,\(^{9–13}\) none of them reported the whole set. The optimized geometries for the different conformers are given in Appendix II.

In the first three conformers the water and sulfuric acid form a six membered ring with water acting both as a donor and an acceptor. As can be seen from Figure 2 the most stable \( \text{g}1 \) geometry can be converted into \( \text{g}3 \) by a rotation of the \( \text{d}_{\text{H}1\text{O}2\text{S}3\text{O}6} \) dihedral angle. It can be thought of as resulting
Figure 2: The four sulfuric acid monohydrate conformers starting from the upper left corner labeled as g1,g2,g3,g4 in order of increasing energy. The angles and bond lengths specified refer to the values in Appendix II from complexation of a g2 sulfuric acid conformer with water. The geometries g1 and g2 are connected by a rotation of the water plane with respect to the sulfuric acid. The fourth conformer differs radically from the first three because, while the six membered ring is still present, water molecule acts solely as a donor and sulfuric acid as an acceptor in this geometry. Additional geometries were also investigated, e.g., a structure with sulfuric acid in g2 conformation in Figure 1 and the water as in g2 in Figure 2, but none of these structures proved to be stable.

The trend of increasing bond angles and decreasing bond lengths when moving to a more complete basis set on the DF-SCS-LMP2 level is continued in Appendix II, with the average change being $-0.025 \text{ Å}$ for the bond length and $0.3^\circ$ for the angles. Similarly, a uniform increase in intramolecular bond lengths is again detected at the CCSD-F12a/CCSD(T)-F12a transition with an average increase of $0.006 \text{ Å}$.

Compared to the experimental geometry by Fiacco et al.,\textsuperscript{35} the CCSD(T)-F12a method yields the best structures with the CCSD-F12a calculations performing approximately as well as the DF-SCS-LMP2/AVTZ calculations and the B3LYP calculation by Bandy and Ianni.\textsuperscript{10} The DF-SCS-LMP2/AVDZ calculations show the largest deviations from the experimental values.
The distortion of the individual sulfuric acid and water molecules upon complex formation is chemically interesting. As can be seen in Figure 2, the g1-g3 conformers of the sulfuric acid monohydrate are held together by an extremely strong hydrogen bond where the sulfuric acid acts as a donor and the water as an acceptor. In the g1 conformer, the length of this hydrogen bond is 1.677 Å at the CCSD(T)-F12a level and it increases somewhat when moving up in energy from g1 to g3. From the length of the hydrogen bond, it is possible to estimate the bond enthalpy. For this hydrogen bond, we obtain an enthalpy of $\Delta H_{m^7} = 31$ kJ mol$^{-1}$.

In the second hydrogen bond ($r_{10}$) of g1, water acts as a donor to the doubly bonded oxygen of sulfuric acid. With the CCSD(T)-F12a method the bond length of this hydrogen bond is 2.140 Å in the g1 conformer. While not as strong as the $r_7$ bond, it is nevertheless important in providing rigidity to the monohydrate molecule. Curiously, $r_{10}$ is at its shortest in the g3 geometry with the CCSD(T)-F12a value of 2.122 Å, where the $r_7$ distance on the other hand has the highest value. In the g1 conformer, the bond enthalpy is estimated to be around $\Delta H_{r7} = 15$ kJ mol$^{-1}$.

The intermolecular distortions are naturally largest in the atoms nearest to the hydrogen bonds. On the CCSD(T)-F12a level, the $r_7$ hydrogen bond results in the lengthening of the $r_6$ OH bond by 0.028 Å in the most stable conformer g1. This together with the surprising strength of the $r_7$ bond can be thought to reflect the earliest states of proton transfer between the two molecules even though the charge-separated complex is not yet stable with only a single water molecule.\cite{10,12,13}

The stretched $r_6$ bond causes a decrease in the $r_5$ bond value of −0.025 Å and an increase in the $a_6$ bond angle by 0.3°. The second hydrogen bond has a smaller impact on the structure with only a 0.009 Å increase in the $r_4$ value and a 0.007 Å increase in the water $r_8$ value in the CCSD(T)-F12a geometries. As a result of the two hydrogen bonds, the water angle is larger by 1.7°. Similar changes are seen in the g2 and g3 conformers. Due to the stronger $r_{10}$ bond in g3, however, the $r_4$ bond is 0.017 Å longer. The results for the other correlation methods agree with the remarks made above.

In the fourth observed geometry with the water molecule located on the top of the sulfuric acid, the two hydrogen bonds are significantly weaker than in all other geometries, which can be seen from the respective bond lengths 2.396 and 2.386 Å at the CCSD(T)-F12a level. From Figure 2 it can be seen that a 180° degrees rotation with respect to the rotational axis that points through the oxygen in the water molecule and the sulfur in sulfuric acid results in a conformation that is indistinguishable from the original one, and so it follows that the point group of this geometry is C$_2$.

Because of the weaker hydrogen bonding present in the g4 conformer, the intramolecular distortion is also smaller. According to the CCSD(T)-F12a calculations, both $r_3$ and $r_6$ bonds are changed by 0.002 Å and angles $a_4$ and $a_5$ are changed by approximately minus one degree.

The energies of the different conformers are shown in Table 2. Again, one sees that with improvements in the correlation treatment, the total energies become lower and the hydrogen bonds decrease in length. In addition, Table 2 shows the interesting fact that while giving the difference between the g1 and g3 levels well, the DF-SCS-LMP2/AVTZ actually predicts that the g2 state lies lower in energy than the g1 state. In my opinion this is because the
Table 2: Energies and relative energies for the geometry optimized structures for the different sulfuric acid monohydrate conformers. The E values are reported in hartree and ∆E values in kJ mol\(^{-1}\).

<table>
<thead>
<tr>
<th></th>
<th>DF-SCS-LMP2/AVDZ</th>
<th>DF-SCS-LMP2/AVTZ</th>
<th>CCSD(T)-F12a/VDZ</th>
<th>CCSD-F12a/VDZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E</td>
<td>∆E</td>
<td>E</td>
<td>∆E</td>
</tr>
<tr>
<td>g1</td>
<td>-775.33076</td>
<td>-775.75353</td>
<td>-775.97921</td>
<td>-775.93093</td>
</tr>
<tr>
<td>g2</td>
<td>-775.33015 1.61</td>
<td>-775.75627 -7.19</td>
<td>-775.97905 0.41</td>
<td>-775.92901</td>
</tr>
<tr>
<td>g3</td>
<td>-775.32895 4.75</td>
<td>-775.75163 4.99</td>
<td>-775.97732 4.96</td>
<td>-775.92901</td>
</tr>
<tr>
<td>g4</td>
<td>-775.31837 32.53</td>
<td>-775.74369 25.82</td>
<td>-775.96493 37.49</td>
<td></td>
</tr>
</tbody>
</table>

domains used in the calculation are too small for accurate energy difference calculations and the fact that the smaller AVDZ basis set gives the right order supports this finding. However, because increasing the domain size slows down the calculations and has only a small effect on near-equilibrium properties such as equilibrium geometries and vibrational frequencies, I opted in general not to increase the domain size. This decision was supported by test calculations with larger domains. For the sulfuric acid monohydrate g1 geometry, increasing the domain size was found to have no effect on the optimized geometry and changed all vibrational wavenumbers by less than 1 cm\(^{-1}\).

With the more accurate explicitly correlated methods, the g1 geometry was found most stable, but is only 0.41 kJ mol\(^{-1}\) below the g2 conformer in energy. The difference between g1 and g3 is found to be 4.96 kJ mol\(^{-1}\). Although not shown in Table 2, the CCSD-F12a energy difference between these two states is found to be 5.04 kJ mol\(^{-1}\).

All the different correlation methods show that the g4 geometry is much higher in energy than any of the other conformations. This means that even though the state exists, its population at atmospheric temperatures is so small that it has no effect on the equilibrium constant. This claim can be verified by comparing the relative Boltzmann populations between the geometries g3 and g4 at the average tropospheric temperature of 268 K.\(^{28}\) From this calculation, one obtains the population ratio g4/g3= 5 \(\times\) 10\(^{-7}\), which shows g4 to be insignificant because the effect of the zero point energies is only about 10 kJ mol\(^{-1}\) as will be shown in section 3.5.

### 3.2 Harmonic Vibrational Wavenumbers

Harmonic wavenumbers were calculated for all different conformers at the DF-SCS-LMP2/AVTZ level. For the g1 geometry, the wavenumbers were also calculated at the CCSD-F12a/VDZ-F12 level. Unfortunately, the wavenumber calculations at the CCSD(T)-F12a/VDZ-F12 level proved to be computationally too demanding for the H\(_2\)SO\(_4\)-H\(_2\)O system.

#### 3.2.1 Water and Sulfuric Acid Molecules

As a first application, the harmonic wavenumbers calculated were used to distinguish between true local minima and stationary points. This was pos-
sible because at stationary points the vibrational motion corresponding to the reaction coordinate has an imaginary value. Furthermore, the $T_1$ diagnostic was used to determine whether the single determinantal wave function of the HF methods gave an accurate enough representation of the system so that the application of correlation methods using only a single determinant reference was justified. The $T_1$ diagnostic is defined with the help of the Coupled Cluster singles amplitude vector $t_1$ by

$$T_1 = \frac{|t_1|}{\sqrt{N_e}},$$

where the single determinantal reference is adequate if $T_1 < 0.02$.

The harmonic wavenumbers for the stable sulfuric acid geometries $g_1$ and $g_2$ and water monomer are given in Appendix III. The $T_1$ diagnostics for these geometries were 0.015 for each. We see from the DF-SCS-LMP2/AVTZ calculations that the harmonic wavenumbers for the different conformers change only around ten wavenumbers for all but the symmetric OSOH bend large amplitude motion. Here, the red shift from $g_1$ to $g_2$ is a bit over two hundred wavenumbers (in reciprocal centimeters), which is a lot. Overall, the largest changes are seen in vibrational modes involving the OH bonds, as can be expected from the nature of the change in geometry.

Comparing the DF-SCS-LMP2 calculations with the CCSD(T)-F12a ones and with the anharmonic experimental values, one finds that the DF-SCS-LMP2 values underestimate the experimental ones for the middle range wavenumbers and overestimate them for the high wavenumbers. Because the harmonic approximation tends to overestimate the wavenumbers, this behaviour implies that for the middle range wavenumbers the DF-SCS-LMP2/AVTZ method tends to underestimate the true wavenumbers. The CCSD(T)-F12a values tend to overestimate the experimental values everywhere but are, in general, of similar quality with the DF-SCS-LMP2 values. In addition to the harmonic approximation, the reason for the overestimation might be that the CCSD(T) method has a habit of overestimating the vibrational harmonic wavenumbers. Normally, this is compensated by basis set incompleteness error, which is greatly diminished in the CCSD(T)-F12 method. Ignoring the OH stretches, the root mean square deviations between the experimental values and those predicted by the DF-SCS-LMP2 and CCSD(T)-F12 methods are 34 and 35 cm$^{-1}$, respectively. The CCSD-F12a calculations habitually overestimate the CCSD(T)-F12a values by more than ten wavenumbers and are thus clearly of lower quality.

The problematic OH stretches are overestimated with both CCSD(T)-F12 and DF-SCS-LMP2 methods by more than 150 cm$^{-1}$ mainly due to the harmonic approximation. As can be seen from the results by Nadykto et al. and Natsheh et al., DFT methods suffer no such drawback, but tend to show worse performance for the middle range frequencies. It was suggested by Miller et al. that the success with the high frequencies is a result of the cancellation of errors between the overly soft DFT frequencies and the increases in frequency caused by the harmonic approximation. The bad performance can be seen from the fact that ignoring the OH stretches, their RMS deviations values are around seventy wavenumbers. Because the middle range vibrational wavenumbers have a higher impact on the vibrational wave function, and because the high-frequency OH-stretches are often easier to treat anharmonically, owing to the fact that they are less coupled to the
other motions, a method that gives good predictions in the middle range of frequencies is preferred. Compared to the DF-SCS-LMP2 and CCSD(T)-F12a values, the DFT methods almost consistently predict lower harmonic wavenumbers.

Because the anharmonic effects are ignored in the harmonic approximation, the harmonic wavenumbers are not completely comparable to the experimental ones. However, as long as the anharmonic effects can be assumed to be small, as is the case in the middle range frequencies, experimental data can be used for the selection of the superior method. To investigate the applicability of the harmonic approximation, Appendix III also lists anharmonic fundamental wavenumbers calculated at the CC-VSCF/MP2-TZP level by Miller et al. For the high frequency OH stretches, these are closer to the experimental ones than any of the \textit{ab initio} results. Surprisingly, for the middle range of frequencies, these show large deviations from the experimental data and have an RMS value of 40 cm$^{-1}$, which is larger than the 34 cm$^{-1}$ obtained for the results of the DF-SCS-LMP2 method. The problems with the anharmonic method are associated with the accurate prediction of the $S-(OH)_2$ stretches and the $S=O_2$ symmetric stretch.

In the case of the water molecule, the stable geometry had a $T_1$ value of 0.008. We can see that results extremely close to the experimental harmonic results can be produced with the CCSD(T)-F12a/VDZ-F12 method. With this approach, the difference between the harmonic and experimental wavenumber is at most four wavenumbers. The DF-SCS-LMP2 method performs in a similar manner as it did in the sulfuric acid case with the RMS deviation being 13 cm$^{-1}$. As before, the DF-SCS-LMP2 values are in general smaller than the CCSD(T)-F12a ones, and the CCSD-F12a values are always larger than the CCSD(T)-F12a ones.

Compared with the results by Kim et al., we see that their CCSD(T) calculation using the nonstandard basis set gives results that agree even better with the experimental values than the CCSD(T)-F12a/VDZ-F12 calculation. On the other hand, both the CCSD(T)/DZP and the regular MP2/AVTZ treatment give results that are significantly worse than the results of the DF-SCS-LMP2 method, with RMS deviations 63 and 23 cm$^{-1}$ respectively.

3.2.2 Sulfuric Acid Monohydrate

Harmonic wavenumbers for the different stable sulfuric acid conformers are given in Appendix IV. Because of its low impact on the equilibrium constant, the structurally different g4 conformer has been omitted from the table. For the presented three structures, the $T_1$ diagnostic values were all 0.014, well under the 0.02 limit.

By comparing the different geometries, we see that the g1 and g3 conformers have vibrational wavenumbers mostly within ten wavenumbers of each other, and the largest differences can be found on the different HOSO bends and the stretching of the bound sulfuric acid OH group. The differences are larger between the g1 and g2 conformers where, for example, many of the intermolecular large amplitude motions like the bound HOSO bend and the S-OH stretch differ by tens of wavenumbers.
The greatest difference between the harmonic spectrum of the bound complex and the spectra of its free constituents is the remarkable redshift of 448 cm that has occurred in the hydrogen bonded OH group. These kinds of large redshifts have been known from experiments to accompany strong hydrogen bonds, and it is possible to use them directly to estimate the bond enthalpy. In the case of the strong hydrogen bond, the estimated bond enthalpy is $\Delta H_r = 28 \text{ kJ mol}^{-1}$, which is a little lower than that predicted from the bond length. The largest changes are observed in the vibrational modes that have changed as a result of the complexations, such as the SOH bends. For the unchanged modes, the shifts are on the average twenty wavenumbers.

Comparison of harmonic wavenumbers with the experimental data is problematic in the sulfuric acid monohydrate case, because the large amplitude motions have significant anharmonicity and all experimental data have been obtained from matrix isolation studies in argon matrices. The system is difficult to study in a precise manner both experimentally and computationally, and a lot more variation can be seen in the experimental values. For the less anharmonic middle range frequencies, once again the DF-SCS-LMP2 calculation gives vibrational wavenumbers that are lower than the CCSD-F12a ones on the average by 27 cm$^{-1}$, with the sole exception being the lowest intermolecular mode. As will be shown in the next section, this mode is so anharmonic that the harmonic values are useless.

Ignoring once again the high-frequency OH stretches, the root mean square deviation between the DF-SCS-LMP2 values and the middle range vibrational wavenumbers of Givan et al. is 25 cm$^{-1}$. Of the whole group, the DF-SCS-LMP2/AVTZ calculation gives the best results for the intermediate wavenumbers. It seems that the DF-SCS-LMP2 method does not systematically over- or underestimate the experimental values, which is unfortunate regarding the possibility of devising appropriate scaling factors. Of the DFT methods, the best performance with respect to the experimental values is obtained with the PW91/TZP method, where the RMS deviation is 37 cm$^{-1}$.

Compared to the DFT PW96/TZP calculations performed by Natsheh et al. for all three geometries considered, the DF-SCS-LMP2 vibrational wavenumbers are lower on average by 36 cm$^{-1}$ for the large amplitude motions. For the high frequency motions, the DF-SCS-LMP2 again overshoot the experimental values by more than a hundred wavenumbers, while the different DFT methods of Nadykto et al. and Natsheh et al. give values at most only few tens of wavenumbers away from the experimental ones. It should be stressed that because the harmonic approximation overestimates the vibrational wavenumbers the results show that the DFT wavenumbers in themselves are not a good representation of the true wavenumbers.

The anharmonicity of the large amplitude intermolecular modes can be seen in the CC-VSCF/Mp2-TZP results of Miller et al. where differences of at least 50 cm$^{-1}$ with the harmonic values are common. As before, the anharmonic calculation yields far better results than any of the other ab initio methods for the OH stretches, but again the agreement with the middle range frequencies is worse than with the DF-SCS-LMP2 with a RMS deviation of 35 cm$^{-1}$. The largest errors between the anharmonic and experimental wavenumbers at these regions are observed at the asymmetric S=O$_2$ stretch.
3.3 Anharmonic Vibrational Wavenumbers

As we have seen in section 2.3, the two factors to have the largest effect on the vibrational partition function are the large amplitude and low frequency vibrations on one hand, and the high frequency vibrations which have the largest effect on the zero point energy on the other hand. Because as a first step, the complete anharmonic treatment would require the calculations of a vast multidimensional PESs, this alone makes it impossible to do on methods such as CCSD(T)-F12a/VDZ-F12, where a single point calculation takes about four hours.

As already mentioned in section 2.2.7, the problems associated with the anharmonic treatment of large molecules were tackled by reserving the anharmonic treatment only to some distinct motions of the system, which on the basis of the harmonic mode analysis could be reasonably isolated from the rest of the system. In this anharmonic domain approximation, two such subsystems were defined for the sulfuric acid monohydrate: The first one was a two-dimensional space consisting of the wagging motion of the free water hydrogen, which connects the conformers g1 and g2, and the free HOSO diedric bend which connects conformers g1 and g3 (for details see section 3.3.2). The second one was a three-dimensional space spanned by the internal coordinates of the two stretches and the bend on the doubly hydrogen-bonded water molecule. To achieve a balanced description of the OH stretches and the HOH bend in the monohydrate complex and free water, also the water monomer was treated anharmonically.

3.3.1 The Three-Dimensional H$_2$O Problem

To obtain the parameters present in equation (103), single point calculations were performed at the CCSD(T)-F12a/VDZ-F12 level around the equilibrium structure by varying one or two of the three water coordinates and keeping all the other coordinates fixed to their equilibrium values. The three two-dimensional surfaces were obtained by varying the bond lengths from −0.375 to 0.338 Å and the angle from −20° to 20° around the equilibrium geometry. Each two-dimensional surface consisted of 169 points. Additional points were calculated for the one-dimensional cuts of the potential energy surfaces, and in the end these consisted of 25 points for the stretches ranging from −0.375 to 0.4875 Å, and 20 points for the bend with the same −20° to 20° range as in the two-dimensional surfaces.

After the surface calculations were complete, the parameters were obtained by least-squares fitting of the multidimensional Taylor series of equation (102) to the surface point data. In order to obtain the best possible fit near the equilibrium geometry, this was done first by finding the one-dimensional parameters of equation (102). In the case of the hydrogen bonded hydrogen atom, the best fit was obtained with a function containing the terms from $T_2^{(b)} z^2$ to $T_6^{(b)} z^6$ and $T_8^{(b)} z^8$, while a polynomial with terms from $T_2^{(f)} z^2$ to $T_7^{(f)} z^7$ was found to give the best fit for the freely vibrating hydrogen. For both fits, the Morse parameter $a$ was first estimated by fitting a fourth order polynomial to the one-dimensional data. The bending PES cut proved to be well described by a fourth order polynomial in the displacement coordinate $\theta$. Because the Taylor series of equation (102) is evaluated at equilibrium, the first order terms in all cases were set to zero. Finally, with all the
one-dimensional parameters known, the two-dimensional parameters were obtained by fitting equation (102) to the two-dimensional PESs with the one-dimensional parameters held fixed. All fittings were done with Mathematica.\textsuperscript{153}

For both the Morse oscillator and the harmonic oscillator wavefunctions the maximum value for the quantum numbers was eight. To further limit the number of wave functions in the calculation it was required that the sum of the two quantum numbers $n_b$ and $n_f$ always be less or equal to eight. After the results were obtained, basis set convergence was checked by doing the same calculation with the maximum values set to nine.

In the water monomer case, the potential energy surfaces were calculated as described above and an eighth order polynomial with all the different terms was used for the OH stretches while a fourth order polynomial was used for the bend. Because of the high symmetry of the monomer system, some of the parameters, such as $f_{r(b)r(b)}$, $f_{r(b)r(r)}$, and $f_{r(b)r(r)}$ have same values. The final parameters are shown in Table 3.

Table 3: Potential energy surface parameters in equation (103) for water in the sulfuric acid monohydrate complex H$_2$O(c) and for isolated water H$_2$O(i). All calculations were performed at the CCSD(T)-F12a/VDZ-F12 level.

<table>
<thead>
<tr>
<th></th>
<th>H$_2$O(c)</th>
<th>H$_2$O(i)</th>
<th>H$_2$O(c)</th>
<th>H$_2$O(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{fb}$ (Å$^{-1}$)</td>
<td>1.8999</td>
<td>1.9950</td>
<td>-0.0833</td>
<td>-0.1245</td>
</tr>
<tr>
<td>$T_2$ (aJ)</td>
<td>1.2217</td>
<td>1.0628</td>
<td>-0.2508</td>
<td>0.1354</td>
</tr>
<tr>
<td>$T_3$ (aJ)</td>
<td>-0.3533</td>
<td>-0.1692</td>
<td>0.3757</td>
<td>0.7061</td>
</tr>
<tr>
<td>$T_4$ (aJ)</td>
<td>0.1162</td>
<td>0.0882</td>
<td>0.6999</td>
<td>0.6302</td>
</tr>
<tr>
<td>$T_5$ (aJ)</td>
<td>0.0149</td>
<td>0.0166</td>
<td>-0.6385</td>
<td>-0.7323</td>
</tr>
<tr>
<td>$T_6$ (aJ)</td>
<td>0.0217</td>
<td>-0.0008</td>
<td>-0.6505</td>
<td>-0.6385</td>
</tr>
<tr>
<td>$T_7$ (aJ)</td>
<td>-0.0015</td>
<td>-0.0031</td>
<td>0.1374</td>
<td>0.2665</td>
</tr>
<tr>
<td>$a_{fb}$ (Å$^{-1}$)</td>
<td>1.8852</td>
<td>1.9950</td>
<td>-0.2324</td>
<td>-0.1610</td>
</tr>
<tr>
<td>$T_2$ (aJ)</td>
<td>1.1872</td>
<td>1.0628</td>
<td>-0.4472</td>
<td>-0.3877</td>
</tr>
<tr>
<td>$T_3$ (aJ)</td>
<td>-0.2652</td>
<td>-0.1692</td>
<td>-0.2565</td>
<td>-0.0216</td>
</tr>
<tr>
<td>$T_4$ (aJ)</td>
<td>0.1171</td>
<td>0.0882</td>
<td>0.2429</td>
<td>0.1281</td>
</tr>
<tr>
<td>$T_5$ (aJ)</td>
<td>0.0125</td>
<td>0.0166</td>
<td>-0.3509</td>
<td>-0.0339</td>
</tr>
<tr>
<td>$T_6$ (aJ)</td>
<td>-0.0124</td>
<td>-0.0008</td>
<td>-0.3509</td>
<td>-0.0339</td>
</tr>
<tr>
<td>$T_7$ (aJ)</td>
<td>-0.0150</td>
<td>-0.0031</td>
<td>0.1374</td>
<td>0.2665</td>
</tr>
</tbody>
</table>

Some of the states calculated from these parameters are shown in Appendixes V and VI for the monohydrate and free H$_2$O, respectively. In Appendix V, the subscript f denotes the free r9 bond and the subscript b refers to the bond between the hydrogen bonded H9 and O8. The state of the system is given as a sum of different $|i⟩_f |j⟩_b |k⟩$ terms, where the $i$ is the quantum number for the free Morse oscillator, $j$ is the quantum number for the bound Morse oscillator, and $k$ is the quantum number for the harmonic oscillator describing the bending motion. The percentage behind the term describes its relative importance to the state.

The effect of the hydrogen bond can be clearly seen in the percentage shift from the fifty-fifty distribution between the $|0⟩_f |i⟩_b |0⟩$ and $|0⟩_f |0⟩_b |0⟩$ states to one where one state always dominates the other. So instead of a symmetric or asymmetric stretch, one should rather be talking about the stretches...
of certain bonds. As the quantum numbers increase, the states become more and more mixed and, for example, the fourth bending overtone already has 20% of $|0\rangle_f |0\rangle_b |6\rangle$ character and 69% $|0\rangle_f |0\rangle_b |5\rangle$ character.

As for the free H$_2$O, we see that for the majority of the states, the difference between the experimental$^{42}$ and calculated results is below ten wavenumbers, with a root mean square deviation of 9.1 cm$^{-1}$. Compared with the CCSD(T)/AVQZ calculations by Salmi et al.,$^{131}$ the CCSD(T)-F12a/VDZ-F12 method gives more accurate values for the lowest three vibrational states, while the overall performance of the standard coupled cluster is better with an RMS deviation of 8.2 cm$^{-1}$. It should be noted that the explicitly correlated method gives much better results than the CCSD(T)/AVTZ calculations performed by Salmi et al.,$^{131}$ which had an RMS deviation of 47.5 cm$^{-1}$.

Based on the fundamental wavenumbers and the first vibrational overtones in appendix VI, the variational calculation tends to predict more accurate values for the bending motions than for the stretches, as was the case in the all the calculations by Salmi et al. This has probably to do with the form of the polynomial chosen for the stretches, and a more balanced treatment could possibly be achieved by introducing higher order coupling terms between the stretches and the bend in equation (103). Curiously enough, the CCSD(T)-F12/VDZ-F12 method seems to increase in accuracy in predicting the second, third and fourth stretching overtones and gives much better values than the CCSD(T)/AVQZ method.

The fundamental wavenumbers are again given in Table 4, and to estimate the effect of the argon matrix on the experimental sulfuric acid monohydrate values, both gas phase and argon matrix values for water are shown. For the bend, the gas phase value is larger by around 5 cm$^{-1}$, while for both OH stretches the values are larger by 20 cm$^{-1}$ so the true values for the sulfuric can be expected to be somewhat larger than those given in Table 4. Even when taking this into account the difference between the experimental$^{41,44}$ and anharmonic values is at most around twenty wavenumbers, which will not have any effect on the vibrational partition function. As can be seen from Appendix IV the experimental values reported for the OH stretches by Givan et al.$^{41}$ are almost 60 cm$^{-1}$ larger than the values by Rozenberg and Loewenschuss$^{44}$ shown in the table. Thus, especially if the effect of the argon matrix is to lower the wavenumbers my computational results strongly support the data by Rozenberg and Loewenschuss. On the other hand, the good quality of our results justifies the approximation that the water molecule can be treated independently while holding the rest of the system rigid.

The two final columns in the table report the shifts in wavenumbers upon complexation for both the anharmonic and experimental values. As is expected, the largest redshift can be seen in the hydrogen bonded hydrogen and the computational result indicates that it is a little smaller than the redshift encountered in the bonded stretching of the water dimer, which according to the results by Salmi et al.$^{131}$ can be predicted to be 97.2 cm$^{-1}$ at CCSD(T)/AVQZ level. According to this, the second hydrogen bond formed in sulfuric acid monohydrate is somewhat weaker than the one formed in water dimer and this is supported by the fact that the bond is also a little shorter with a length of 1.96 Å compared with the 2.140 Å of the second hydrogen bond in the monohydrate. It should be kept in mind, though, that
The fundamental wavenumbers for free water and water bound in the sulfuric acid monohydrate complex at the CCSD(T)-F12a/VDZ-F12 level compared with experimental values. All values are reported in cm\(^{-1}\). The shifts in wavenumber for the anharmonic \(\Delta \tilde{\nu}_a\) and experimental \(\Delta \tilde{\nu}_e\) values are defined as the free water fundamental wavenumber minus the fundamental wavenumber in the hydrate. \(^a\)Engdahl and Nelander\(^{43}\) in Ar matrix, \(^b\)Benedict et al. in the gas phase, \(^c\)Givan et al.\(^{41}\) in Ar matrix, \(^d\)Rozenberg and Loewenschuss\(^{44}\) in Ar matrix.

<table>
<thead>
<tr>
<th>HOH bend</th>
<th>isolated H(_2)O</th>
<th>bound H(_2)O</th>
<th>bound H(_2)O</th>
<th>(\Delta \tilde{\nu}_a)</th>
<th>(\Delta \tilde{\nu}_e)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>anharm.</td>
<td>exptl.(^a)</td>
<td>exptl.(^b)</td>
<td>anharm.</td>
<td>exptl.</td>
</tr>
<tr>
<td>HOH bend</td>
<td>1594.5</td>
<td>1589.1</td>
<td>1594.6</td>
<td>1589.2</td>
<td>1599.7(^c)</td>
</tr>
<tr>
<td>sym. stretch (bonded)</td>
<td>3651.9</td>
<td>3638.0</td>
<td>3656.7</td>
<td>3579.7</td>
<td>3581.9(^d)</td>
</tr>
<tr>
<td>asym. stretch (free)</td>
<td>3761.0</td>
<td>3734.3</td>
<td>3755.8</td>
<td>3725.7</td>
<td>3696.6(^d)</td>
</tr>
</tbody>
</table>

the hydrate structure is a lot more rigid than the water dimer due to the presence of the stronger hydrogen bond \(r_7\) in Figure 2.

If the effect of the argon matrix has an approximately constant effect on the vibrational wavenumbers, then the shifts between the anharmonic and experimental values should be comparable. With regard to the shifts in wavenumbers, the best description is achieved for the free OH. This is logical, as this degree of freedom is probably the least coupled to the intermolecular modes and the vibrational modes in the sulfuric acid. Because of the hydrogen bond, the strongest couplings are most likely found with the hydrogen bonded OH stretch. Table 4 confirms that this is the case. Because the HOH bend has the effect of bending the hydrogen bonded hydrogen towards the sulfuric acid molecule, rather strong couplings can be expected for this degree of freedom as well – a notion that is supported by the 16 cm\(^{-1}\) difference between the anharmonic and experimental shifts in wavenumber.

### 3.3.2 The Two-Dimensional Large Amplitude Problem

As described in section 2.2.7, the two-dimensional PES consisted of the torsional rotation of H1 hydrogen around the S3-O2-axis and the wagging motion of the H10 hydrogen. These two motions connect the geometries g1, g2 and g3, and so their anharmonic treatment is of fundamental importance if an accurate value for the equilibrium constant is to be obtained.

The two-dimensional PES was determined on the CCSD(T)-F12a/VDZ-F12 level and consisted of a total of 117 points. The torsional angle \(\tau\) was allowed to vary from \(-150^\circ\) to \(180^\circ\) and the wagging angle \(\theta\) from \(-90^\circ\) to \(90^\circ\) with the average increment being \(30^\circ\) for both angles. As in the three-dimensional case, the parameters for equation (109) were obtained first by doing two separate least-squares fittings with the eighth order polynomial and Fourier series to one dimensional-cuts of the surface with the other variable held at its equilibrium value. For the torsional motion, the cut had 36 points, and for the wagging angle, the number of points was 21. Because the geometry optimizations at the CCSD(T)-F12a level are very demanding computationally, the structure could not be allowed to relax.
at each point. Instead, for the one-dimensional wagging calculations the geometry was frozen to an average between the two minimum conformers g1 and g2. According to test calculations, while this approximation did raise the energies of both the g1 and the g2 states somewhat, it introduced only a minimal error in the energy difference between the two states. For the torsional motion, a linear mapping was introduced to stepwise change the geometry from the average one used in the wagging calculation to the g3 one.

Several different forms for the potential energy surface were considered but the one in equation (109) was chosen because it gave the best results for the one-dimensional fits. The two-dimensional calculations were performed by V. Hänninen except for some preliminary one-dimensional calculations for the torsional rotation with the DF-SCS-LMP2/AVTZ method, which were performed by the author.

The parameter values obtained from the potential surface fitting are shown in Table 5 and the fitted potential energy surface can be seen in Figure 3. The three identified minima can be clearly distinguished with the lowest g1 conformation at the front of the picture and the g2 and g3 conformers corresponding to the minima on the right and the left, respectively.

### Table 5: Potential energy surface parameters in cm$^{-1}$

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0$</td>
<td>1056.67</td>
<td>$C_2$</td>
<td>-834.58</td>
</tr>
<tr>
<td>$A_1$</td>
<td>-90.70</td>
<td>$C_3$</td>
<td>28.35</td>
</tr>
<tr>
<td>$A_2$</td>
<td>458.43</td>
<td>$C_4$</td>
<td>533.63</td>
</tr>
<tr>
<td>$A_3$</td>
<td>-71.39</td>
<td>$C_5$</td>
<td>-11.12</td>
</tr>
<tr>
<td>$A_4$</td>
<td>5.59</td>
<td>$C_6$</td>
<td>-71.05</td>
</tr>
<tr>
<td>$A_5$</td>
<td>4.27</td>
<td>$C_7$</td>
<td>2.21</td>
</tr>
<tr>
<td>$A_6$</td>
<td>0.14</td>
<td>$C_8$</td>
<td>7.57</td>
</tr>
<tr>
<td>$B_1$</td>
<td>-218.65</td>
<td>$f_{τφ}$</td>
<td>-6.16</td>
</tr>
<tr>
<td>$B_2$</td>
<td>20.60</td>
<td>$f_{ττφ}$</td>
<td>12.10</td>
</tr>
<tr>
<td>$B_3$</td>
<td>-15.26</td>
<td>$f_{τφφ}$</td>
<td>0.52</td>
</tr>
<tr>
<td>$B_4$</td>
<td>4.09</td>
<td>$f_{τττφ}$</td>
<td>10.66</td>
</tr>
<tr>
<td>$B_5$</td>
<td>-0.48</td>
<td>$f_{ττττ}$</td>
<td>4.08</td>
</tr>
<tr>
<td>$B_6$</td>
<td>0.07</td>
<td>$f_{ττττ}$</td>
<td>6.44</td>
</tr>
</tbody>
</table>

The matrix elements required in equation (112) were calculated numerically with Mathematica using adaptive algorithms and letting the program automatically choose the best method for numerical integration. After the diagonalization procedure the eigenvalues of energy could be obtained. The maximum number for the vibrational quantum number was 12 and the rotational quantum number had values $k \in [-11,11]$.

The states below 1000 cm$^{-1}$ are shown in Figure 4 along with the two one-dimensional cuts from Figure 3 obtained by varying only one of the variables and keeping the other fixed to its equilibrium value. All of the states up to 1500 cm$^{-1}$ are also given in Appendix VII along with the one-dimensional
results where all the couplings have been ignored for both the torsional and the wagging motion calculated with the CCSD(T)-F12a method. Finally, the table shows the results for the preliminary DF-SCS-LMP2/AVTZ calculations for the one-dimensional torsional motion.

Figure 3: The fitted potential energy surface for the torsional rotation of H1 and the wagging motion of H10. Both the $\tau$ and $\vartheta$ axes are in radians. Blue curve is the one-dimensional PES cut for the wagging motion and the purple one is the one-dimensional PES cut for the torsional motion.

Figure 4 is interesting in several respects. We see that the barrier height between the g1 and g2 geometries is only 384 cm$^{-1}$, which corresponds to 4.6 kJ mol$^{-1}$. On the other hand, the barrier height between the g1 and g3 geometries is much higher: 996 cm$^{-1}$ or 11.9 kJ mol$^{-1}$, which is similar in size to the torsional barrier in ethane.$^{154}$ We also see that the zero point energy is 297 cm$^{-1}$, which is almost double the value 155 cm$^{-1}$, one would predict from the harmonic approximation. The first transition is 34 cm$^{-1}$, which deviates from the harmonic prediction of 47 cm$^{-1}$ as shown in appendix IV. From Figure 4, we see that this transition arises from the splitting of levels caused by quantum mechanical tunneling. Looking at the next vibrational overtone, a harmonic transition would occur at 94 cm$^{-1}$ whereas the anharmonic picture predicts this value to be 237 cm$^{-1}$. This total failure of the harmonic approximation can be attributed to the multistationary shape of the PES.

It is interesting to note that the fundamental anharmonic transition for the water wagging motion 34 cm$^{-1}$ is also completely different from the fundamental anharmonic transition of 104 cm$^{-1}$ predicted by Miller et al.$^{30}$ at the CC-VSCF/MP2-TZP level. Similar kind of large discrepancies are obtained for the torsion as the values are 259.4 cm$^{-1}$ and 396.5 cm$^{-1}$, respectively. The underlying reason for this disagreement is probably the fact that the CC-VSCF/MP2-TZP calculation has been performed on the g1 conforma-
Figure 4: The lowest sixteen states of the variational calculation. The blue curve is the wag angle PES cut and the purple one refers to the torsional PES cut with the $\vartheta$ angle held fixed. Horizontal axis is in radians.

The tunneling effect is also clearly visible in the splittings of the higher excited states. As can be seen, the density of states increases with increasing wavenumber.

There seems to be a discontinuity in the potential energy surface at the g1 structure because the global minima on the one-dimensional wagging and torsional curves are not on the same level. The reason for this is that the sixth order Fourier series had trouble predicting rather sharp curve in the immediate vicinity of the minimum. The 8.1 cm$^{-1}$ error arising from this effect was corrected in the final results by substracting it from the eighth order polynomial of the torsional motion before the calculation of the matrix elements.

From Appendix VII, we see that the one-dimensional CCSD(T)-F12a calculations predicts a zero point energy that is only 3.5 cm$^{-1}$ from the two-dimensional one. In addition, up to around 750 cm$^{-1}$ the states are all less than ten wavenumbers apart from the corresponding ones in the two-dimensional calculation. The inclusion of the coupling terms only has large impact on some of the states, such as the one at 791.5 cm$^{-1}$ in the two-dimensional picture. Comparison between the one-dimensional torsional motions calculated with the different correlation methods show that, while...
the agreement is good for the first two states, the difference between the third state is 52 cm$^{-1}$. This arises from the fact that the torsional wall has a different shape in the DF-SCS-LMP2/AVTZ calculation, again probably due to the too small size of the local domains in the calculation.

### 3.4 Partition Functions

To get to the thermodynamic properties of the system, the canonical partition function of equation (120) has to be found. This was done by separately calculating the different partition functions in equation (123). As described in section 2.3, the nuclear partition function was ignored.

The harmonic parts of the vibrational partition function were obtained directly from equation (128), while the anharmonic parts were calculated from the definition of the molecular partition function in equation (121) by summing over the different vibrational states until convergence was achieved. The harmonic and anharmonic parts were then multiplied together to form the vibrational partition function. In the case of large amplitude motions, the sum of equation (121) was over 266 states and extended up to 5132 cm$^{-1}$, where the contribution of a single term to the sum was of the order of $10^{-11}$. As for the high frequency calculations on the free and bound water molecules, convergence was achieved much faster and the contribution was less than $10^{-15}$ after only eleven states.

The electronic partition functions were calculated from equation (125), with all the ground states non-degenerate. The non-degeneracy assumption was made because all the electrons in the molecules are paired. The energies $D_e$ were obtained from the CCSD(T)-F12a geometry optimizations. Equation (125) was simplified by ignoring all but the first term because the electronically excited states for all three molecules lie high above the ground state. In the case of the water monomer, both experimental$^{155}$ and computational$^{156}$ results predict that the singlet state excitation energy is around 720 kJ mol$^{-1}$. For the sulfuric acid molecule, computational studies$^{157}$ predict the lowest triplet and singlet states to be 790 and 810 kJ mol$^{-1}$ above the ground state, respectively, and the experimental values support this.$^{149}$ The excitation energies of the sulfuric acid monohydrate have not yet been calculated or experimentally measured, but judging from the sulfuric acid and water results can be assumed to lie within the same region.

Calculation of the rotational partition function from equation (126) requires the knowledge of the rotational constants. These were obtained as a byproduct of the geometry optimizations on the CCSD(T)-F12a/VDZ-F12-level. Due to the $C_2$ symmetry of the lowest state of sulfuric acid and $C_{2v}$ symmetry of the water monomer, these molecules have the symmetry number $\sigma = 2$. The lack of symmetry in the monohydrate complex and the $g_2$ sulfuric acid conformer results in a symmetry number $\sigma = 1$.

In the derivations of equations (124) and (126), an integral approximation has been applied to the definition (121), which assumes that the energy levels in equation (121) are close to one another. Because the rotational energies are linearly dependent on the rotational constants, the approximation is valid when working above the rotational temperature $\Theta$, defined as

$$\Theta = \frac{Ach}{k},$$  \hspace{1cm} (141)
where $A$ is a rotational constant. For the translational degrees of freedom, function (124) is applicable when the size of the system is large enough so that the energies in the particle-in-a-box model lie in close proximity. This is true when the thermal wavelength $\Lambda$, defined by

$$\Lambda = \frac{\hbar}{(2\pi mkT)^{1/2}}$$ (142)

is much smaller than the average separation between the particles in the system.

Rotational constants and the resulting rotational partition functions are given in Table 6, which also shows the values for the standard molar translational partition function $q_{m,t}$ from equation (124) with $V_m = RT/p$. To assess the reliability of equations (124) and (126), Table 6 also lists rotational temperatures and thermal wavelengths.

### Table 6: Calculated and experimental rotational constants $A$, rotational temperatures $\Theta$, rotational partition functions $q_r$, thermal wavelengths $\Lambda$ and standard molar translational partition functions. $^a$Cook, $^b$Hintze et al., $^c$Fiacco et al.

<table>
<thead>
<tr>
<th></th>
<th>$A$ / cm$^{-1}$</th>
<th>$\Theta$ / K</th>
<th>$q_r$</th>
<th>$\Lambda$ / Å</th>
<th>$q_{m,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>calc.</td>
<td>exp(t)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_2O$</td>
<td>27.211</td>
<td>27.881$^a$</td>
<td>39.2</td>
<td>0.238</td>
<td>1.8 $\cdot$ 10$^{30}$</td>
</tr>
<tr>
<td></td>
<td>14.589</td>
<td>14.522$^a$</td>
<td>21.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.497</td>
<td>9.278$^a$</td>
<td>13.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_2SO_4$</td>
<td>0.172</td>
<td>0.172$^b$</td>
<td>0.25</td>
<td>3.9 $\cdot$ 10$^4$</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>0.169</td>
<td>0.167$^b$</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.162</td>
<td>0.163$^b$</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_2SO_4$-H$_2O$</td>
<td>0.169</td>
<td>0.24</td>
<td>2.0 $\cdot$ 10$^5$</td>
<td>0.094</td>
<td>3.0 $\cdot$ 10$^{31}$</td>
</tr>
<tr>
<td></td>
<td>0.064</td>
<td>0.063$^c$</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.063</td>
<td>0.063$^c$</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because tropospheric temperatures vary roughly from 210 to 300 K, we see that $T \gg \Theta$ and the rotational partition function is perfectly applicable. As for the translational partition function, the average separation between the particles can be calculated from the ideal gas law and is found to be 30 Å at $p^0 = 1$ bar with $T$ between 210 and 300 K. Comparison with the $\Lambda$ values in the table shows that the approximation is also valid in this case.

Comparison of the calculated rotational constants with the experimental ones shows that the agreement is not very good when taking into account the fact that rotational constants can be determined with the accuracy of eight significant numbers. On the other hand, the computational results are not directly comparable to the experimental ones because the experimental constants are determined at the ground vibrational state while the computational rotational constants correspond to equilibrium values. Luckily, the rotational partition function and especially the thermodynamic properties are not too strongly dependent on the rotational constants, and so the agreement can be considered as satisfactory.
In general, the partition function can be viewed as a rough estimate of the number of thermally accessible states of the molecule. Applying this interpretation to the partition functions in Table 6, we see that the number of available rotational levels is large at room temperature, but is nowhere near the number of occupied translational levels. Judging from the \( q_t \) values, the total number of states thermally accessible to the system is extremely large, and so the approximation required for the separation in equation (119) is considered valid.

Unfortunately, the price for the theoretical rigor obtained by our choice of the zero-point energy is the fact that the physical interpretation of the partition function is lost on both the electronic and vibrational parts. This is the reason their values are not shown in Table 2.3.

### 3.5 Thermodynamic Parameters

After the partition functions were calculated, a substitution of these into equation (132) gives the equilibrium constant. At equilibrium, the easiest way to calculate the standard Gibbs free energy is using the fundamental equation

\[
-RT \ln K = \Delta_r G^\circ_m.
\]  

The standard molar reaction enthalpies can be calculated from equation (135) and after this, the standard molar reaction entropy can be obtained directly from the definition of the Gibbs free energy

\[
G = H - TS.
\]

The thermodynamic data for the three most significant reaction paths, with purely harmonic vibrational frequencies calculated at the DF-SCS-LMP2/AVTZ level, are shown in Table 7. Surprisingly, the effect of the zero point energies is to make the reaction leading to the g2 conformer most exoergic, followed by the reaction with the g2 sulfuric acid and water to form the g3 conformer of the monohydrate. The g3 reaction is seen to be the most exothermic with a value of \(-47.0 \text{ kJ mol}^{-1}\), while the g2 reaction is the least exothermic with \(\Delta_r H^\circ_m = -45.7 \text{ kJ mol}^{-1}\). The standard molar entropy is largest for the reaction with g2 as a product and smallest for the reaction with g3 as a product. Because the effect of the negative entropy term is to increase the reaction Gibbs energy linearly with increasing temperature, it follows that the equilibrium constant is largest for the g2 reaction, in accordance with equation (143). This fundamentally flawed result is a consequence of the failure of the harmonic approximation. As can be seen from Appendix IV, the first three normal modes have much smaller harmonic wavenumbers for the g2 conformer compared with the g1 values. Through equation (128), these low energy vibrations have the largest effect on the vibrational partition function by providing many energetically accessible states. The ultimate result is that the vibrational partition function for the g2 sulfuric acid monohydrate conformer has a value seven times higher than that of the g1 conformer.

From a qualitative point of view, the thermodynamic parameters are reasonable. The enthalpies have a negative value because the formation of the two hydrogen bonds releases heat. Entropy is negative because the hydrogen bonds limit the movement of the individual molecules and the reaction results in a single organized structure. Due to the strong hydrogen bonding
Table 7: Standard molar reaction energies, enthalpies, entropies, Gibbs free energies and equilibrium constants for the reactions with H$_2$SO$_4$ in the g1 conformation and H$_2$SO$_4$·H$_2$O in the g1 conformation (R1), with H$_2$SO$_4$ in the g1 conformation and H$_2$SO$_4$·H$_2$O in the g2 conformation (R2), and with H$_2$SO$_4$ in the g2 conformation and H$_2$SO$_4$·H$_2$O in the g3 conformation (R3). All values are calculated using the harmonic approximation.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta_r E_m^0$ / (kJ mol$^{-1}$)</th>
<th>$\Delta_r H_m^0$ / (kJ mol$^{-1}$)</th>
<th>$\Delta_r S_m^0$ / (J mol$^{-1}$ K$^{-1}$)</th>
<th>$\Delta_r G_m^0$ / (kJ mol$^{-1}$)</th>
<th>$K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>-43.9</td>
<td>-46.1</td>
<td>-117.8</td>
<td>-11.0</td>
<td>83.2</td>
</tr>
<tr>
<td>R2</td>
<td>-44.4</td>
<td>-45.7</td>
<td>-99.2</td>
<td>-16.1</td>
<td>654.9</td>
</tr>
<tr>
<td>R3</td>
<td>-44.1</td>
<td>-47.0</td>
<td>-132.7</td>
<td>-7.5</td>
<td>20.5</td>
</tr>
</tbody>
</table>

present in the complexes, the standard molar Gibbs energy is negative, so at least at 298.15 K and 1 bar the complexes are stable.

Thermodynamic data with the anharmonic domain approximation incorporated into the calculation are shown in Table 8. The differences in the reaction energies are due to the differences in zero point energies. It can be seen that the result of the anharmonic corrections is to increase the zero point energy as was evident already from Figure 4 and Appendix VII. The standard molar reaction enthalpies are mostly determined by the reaction energies, so the rather uniform distribution in Table 8 can be expected. As for the entropies, the CCSD-F12a/VDZ-F12 without anharmonic domains has a $\Delta_r S_m^0$ value 33.7 J mol$^{-1}$ K$^{-1}$ higher than the CCSD-F12a/VDZ-F12 calculation containing them. The difference is 4.5 J mol$^{-1}$ K$^{-1}$ with the DF-SCS-LMP2 methods. From Appendix IV, we see that the CCSD-F12a/VDZ calculation has a lowest harmonic vibrational wavenumber of 1 cm$^{-1}$ which results in a huge vibrational partition function, high entropy, and a very small Gibbs free energy. This very low vibrational frequency arises from the fact that the CCSD-F12a/VDZ-F12 potential energy surface is rather flat around the equilibrium geometry. Because the application of a higher level computational treatment for the calculation of electronic energies is relatively straightforward, my findings support the commonly-held belief that it is the accurate evaluation of the reaction entropy that is the main problem in estimating the reaction Gibbs energies.

There is a striking similarity between the DF-SCS-LMP2/AVTZ and CCSD-F12a/VDZ-F12 calculations with the anharmonic domains. The energies and enthalpies differ by 0.1 kJ mol$^{-1}$, the ZPE are virtually the same and the reaction entropies differ only by 0.8 J mol$^{-1}$ K$^{-1}$. All these factors lead to a Gibbs energy difference of only 0.2 kJ mol$^{-1}$ between the different treatments. I interpret this result so that the remaining errors in the calculation do not arise from a particular *ab initio* harmonic treatment, but from the harmonic approximation in general. This is important because we already know that the harmonic approximation works badly for the remaining large amplitude motions and the high frequency OH stretches, and so further anharmonic domains can be chosen to cover these. With the addition of anharmonic domains, the results can be rather systematically improved towards the true thermodynamic values.
Table 8: Reaction energies, zero point energies, enthalpies, entropies and Gibbs free energies with the harmonic frequencies calculated using either the DF-SCS-LMP2/AVTZ method or the CCSD-F12a/VDZ-F12 method. The calculations involving the anharmonic domain approximation are denoted by the additional (a). $\Delta_r E_m$, $\Delta_r H_m$, $\Delta ZPE$ and $\Delta_r G_m$ values are given in kJ mol$^{-1}$, while $\Delta_r S_m$ values are given in J mol$^{-1}$ K$^{-1}$. All values are reported at 1 atm and 298 K unless otherwise stated. a Hanson and Eisele, b Kürtén et al., c with binding energies calculated at the CCSD(T)/cc-pV(T+d)Z level, d Kürtén et al. with a three step method and anharmonic corrections, e Al Natsheh et al. at 298.15 K, f Re et al., g Bandy and Ianni, and h with anharmonic corrections. Except for the calculations involving anharmonic domains, all computational results correspond to the reaction with H$_2$SO$_4$ in the g1 conformation and H$_2$SO$_4$·H$_2$O in the g1 conformation.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\Delta_r E_m$</th>
<th>$\Delta ZPE$</th>
<th>$\Delta_r H_m$</th>
<th>$\Delta_r S_m$</th>
<th>$\Delta_r G_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF-SCS-LMP2/AVTZ</td>
<td>-43.9</td>
<td>9.1</td>
<td>-46.1</td>
<td>-117.7</td>
<td>-11.0</td>
</tr>
<tr>
<td>DF-SCS-LMP2/AVTZ (a)</td>
<td>-42.5</td>
<td>10.6</td>
<td>-44.9</td>
<td>-123.2</td>
<td>-8.2</td>
</tr>
<tr>
<td>CCSD-F12a/VDZ-F12</td>
<td>-44.4</td>
<td>8.7</td>
<td>-46.3</td>
<td>-88.3</td>
<td>-20.0</td>
</tr>
<tr>
<td>CCSD-F12a/VDZ-F12 (a)</td>
<td>-42.4</td>
<td>10.6</td>
<td>-45.0</td>
<td>-124.0</td>
<td>-8.0</td>
</tr>
<tr>
<td>exptl. a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-15.1</td>
</tr>
<tr>
<td>MP2/6-311++G(2d,2p)</td>
<td>-35.5</td>
<td>9.2</td>
<td>-37.7</td>
<td>-118.7</td>
<td>-2.4</td>
</tr>
<tr>
<td>MP2/cc-pV(T+d)Z</td>
<td>-39.4</td>
<td>9.4</td>
<td>-41.9</td>
<td>-120.9</td>
<td>-5.9</td>
</tr>
<tr>
<td>MP2/cc-pV(T+d)Z</td>
<td>-38.5</td>
<td>9.4</td>
<td>-49.8</td>
<td>-123.3</td>
<td>-14.4</td>
</tr>
<tr>
<td>3-step d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PW91/TZP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PW91/cc-pV(T+d)Z</td>
<td>-45.0</td>
<td>9.3</td>
<td>-48.2</td>
<td>-126.3</td>
<td>-10.5</td>
</tr>
<tr>
<td>B3LYP/D95++(d,p) f</td>
<td>-43.5</td>
<td>10.0</td>
<td>-46.4</td>
<td>-122.6</td>
<td>-10.0</td>
</tr>
<tr>
<td>B3LYP/6-311++G(2d,2p) g</td>
<td>9.6</td>
<td>40.6</td>
<td>-128.0</td>
<td>-2.5</td>
<td></td>
</tr>
<tr>
<td>B3LYP/cc-pV(T+d)Z</td>
<td>9.6</td>
<td>49.8</td>
<td>-123.3</td>
<td>-13.1</td>
<td></td>
</tr>
<tr>
<td>B3LYP/cc-pV(T+d)Z</td>
<td>9.0</td>
<td>49.9</td>
<td>-117.8</td>
<td>-14.8</td>
<td></td>
</tr>
</tbody>
</table>
Of the other computational studies, only Kürten et al. corrected for the BSSE with the counterpoise correction, and the table shows the CP corrected values of their results for all except the B3LYP/cc-pV(T+d)Z method. They also found that the correction was necessary with all but the largest aug-cc-pV(T+d)Z basis set they employed. Conversely, a previous study by Beichert and Schrems at the MP2/6-311++G(2d,2p) showed the BSSE to be negligible for that method. The large difference between the values provided by al Natsheh and Bandy and Ianni was explained by Kürten et al. to result from the fact that the PW91/TZP method predicts very high stabilities whereas the B3LYP/6311++G(2d,2p) has a tendency to yield very low stabilities. In general, it can be seen that the choice of the method and basis used has a large effect on the results. It should be noted that excluding our results that use the anharmonic domain approximation, all the thermodynamic parameters reported in Table 8 correspond to the reaction of the g1 sulfuric acid with water to yield the g1 conformer of the sulfuric acid monohydrate.

For all except the methods containing anharmonic domains, the zero point energies are generally within 1 kJ mol$^{-1}$ of each other, so my results further confirm the conclusion made by Kürten et al. on the basis of their study that it is justified to treat the vibrational properties at a lower level of theory. As the inclusion of anharmonic domains leads to an increase of about 1.5 kJ mol$^{-1}$, I would like to add that even if a lower level treatment is used, some of the degrees of freedom must be treated anharmonically. In this light, it would seem that the largest source of differences in the past results arise from the inaccurate calculation of the electronic energies.

An interesting comparison can be made between our results and the MP2/cc-pV(T+d)Z calculation where the binding energies were calculated at the CCSD(T)/cc-pV(T+d)Z level. The CCSD(T) method seems to predict bonding energies that differ 5.0 kJ mol$^{-1}$ from the CCSD(T)-F12a/VDZ-F12 method, and the reaction energies which include the zero point energies are off by 3.8 kJ mol$^{-1}$ compared with the calculations that include anharmonic domains. The Gibbs energy is around 3.0 kJ mol$^{-1}$ larger for the MP2/cc-pV(T+d)Z method. Comparison with the completely harmonic DF-SCS-LMP2/AVTZ method shows even larger deviations in Gibbs free energy, partly because the zero point energies are very similar. In my opinion the three important reasons for these differences are that the geometry was not optimized at the CCSD(T)/cc-pV(T+d)Z level in the MP2/cc-pV(T+d)Z calculation, the CCSD(T) method itself gives different results and the harmonic approximation was used throughout. As mentioned earlier and shown on several occasions in this investigation, the CCSD(T)-F12a/VDZ-F12 method generally gives results of CCSD(T)/AVQZ quality, furthermore the DF-SCS-LMP2 method is by construction virtually free of the basis set superposition error, so the results represented in my study are more trustworthy.

Kürten et al. also measured the effect of anharmonic corrections at the non-CP corrected B3LYP/cc-pV(T+d)Z level. The effect of anharmonicity can be most clearly seen in the zero point energy, which for the B3LYP/cc-pV(T+d)Z calculations differs by 0.6 kJ mol$^{-1}$. This is much smaller and of a different sign than the change in zero point energies caused by the inclusion of anharmonic domains. The reason for this might be that the corrections are only relatively small modifications to the harmonic frequencies and in-
herently assume that the harmonic approximation is at least somewhat valid, which is not the case in the torsional and wagging motions. The small values for enthalpy and Gibbs free energy in the B3LYP/cc-pV(T+d)Z values are due to BSSE. The CP corrected harmonic enthalpy was $-41.2 \text{ kJ mol}^{-1}$ and the Gibbs energy was $-4.8 \text{ kJ mol}^{-1}$, which are in better agreement with the rest of the results.

The most accurate results to date are probably the ones obtained using a three step method by Kúrten et al.\textsuperscript{159} In their method, the geometry optimization and harmonic frequency calculations were done at the MP2/aug-cc-pV(D+d)Z level. After this an MP2/aug-cc-pV(T+d)Z calculation was used to estimate the complete basis set limit via a two parameter formula and an MP4/aug-cc-pV(D+d)Z calculation was used to incorporate correlation effects beyond the MP2 level. Finally, anharmonicity was taken into account by calculating the anharmonicity constants perturbatively, and then using statistical thermodynamics to calculate the internal energy and entropy from these. The fact that their predicted reaction energies differ by 2 kJ mol$^{-1}$ from our results is reflected in the relatively small enthalpy difference of 1.5 kJ mol$^{-1}$ between the two methods. Likely the main reason for this difference is that the CCSD(T)-F12a method can recover more of the correlation energy than the MP4 method. The large difference in entropies arises probably because their anharmonic treatment does not account for the presence of the different H$_2$SO$_4$·H$_2$O isomers. Due to the higher value of reaction entropy in the 3-step method the value of reaction Gibbs free energy is substantially lower than the one obtained with anharmonic domains.

On the basis of equation (138) one would draw the conclusion that the inclusion of anharmonic effects into the sulfuric acid-monohydrate system would reduce the Gibbs energy. This is because the number of states in the anharmonic system is larger which leads to a larger value for the vibrational partition function. The increase in vibrational partition function is largest for the sulfuric acid monohydrate molecule due to its larger number of vibrational degrees, and so the result would be an overall decrease in reaction Gibbs free energy. This behaviour is confirmed for the B3LYP/cc-pV(T+d)Z calculation but not found in our results. This is because for the torsional and wagging motions studied, the anharmonic treatment leads to an increase in zero point energy accompanied by a radical change in the distribution of states which is not compensated by the creation of new states at higher frequencies. The flush of states from the low frequencies to higher ones when moving from the harmonic approximation to the anharmonic treatment can most clearly be seen from Figure 4. In the harmonic approximation, there are 10 states below 500 cm$^{-1}$ all of which would give a significant contribution to the partition function. Even though the number of states above 500 cm$^{-1}$ is greater in the anharmonic picture than in the harmonic one, the exponential dependence on the energy difference present in equation (121) means that the impact of these states is much smaller.

The reorganization in the density of states is even more severe for the g2 and g3 geometries of the sulfuric acid monohydrate, due to their lower values for the first intermolecular vibrational mode. This is why the values in Table 7 are so much greater than the anharmonic domain value given in Table 8 that accounts for equilibrium constant values with all the g1, g2 and g3 conformers of the monohydrate.
Finally, the experimental Gibbs energy measured by Hanson and Eisele is also displayed in Table 8 and is in closest agreement with the PW91/TZP results by Natsheh et al., the non-CP corrected anharmonic B3LYP/cc-pV(T+d)Z results of Kürtken et al., and the results obtained from the 3-step method by Kürtken et al. In my opinion the experimental value is problematic, mainly because of the large number of different approximations used in its calculation. For example, in their study Hanson and Eisele were forced to estimate the relative importance of the different hydrates and ignored all clusters with more than two water molecules and clusters containing two or more sulfuric acid molecules. This approximation is challenged by the fact that in the computational study at the B3LYP/D95++(d,p) level, Re et al. obtained reaction Gibbs energies for the formation of the \( \text{H}_2\text{SO}_4(\text{H}_2\text{O})_n \) complexes via reaction (19) with \( n = 3 \) or 4 that were similar in size to the \( \Delta_rG_m \) values of the monohydrate formation. Furthermore, the molecular dynamics simulation of Kusaka et al. indicates that the importance of the higher-order clusters is important in hydration. Because of these difficulties, the high-quality computational results represented in this study are probably closer to the truth than the reported experimental value.

### 3.5.1 Temperature and Pressure Dependence of the Thermodynamic Properties

The temperature dependence of the equilibrium constant was calculated from equation (136), which takes into account the temperature dependence of the reaction enthalpy for the reaction of a sulfuric acid in the g1 conformer with water yielding any of the three most stable sulfuric acid monohydrate structures. Because the temperature profile of the troposphere is almost linear, the equilibrium constant can be represented as a function of the elevation from sea level instead of a function of temperature. An equation of the form \( h = \frac{(292 - T)}{6.5} \) was used to calculate the elevation in kilometers from the temperature, and the result is shown in Figure 5. In addition to the equilibrium constants calculated from equation (136), the figure also shows the results when the enthalpy is held fixed at its value at 298 K.

Figure 5 shows clearly the exponential dependence on the temperature of the equilibrium constant as implied in equation (143). Because of the negative values of entropy and enthalpy and their relatively minor temperature dependences, the equilibrium constant decreases with increasing temperature. The constant enthalpy approximation is good down to 230 K, but below this the temperature dependent equation should be used. This means that the temperature dependence of the reaction enthalpy is only important around the tropopause and can be ignored for most of the stratosphere and troposphere.

In addition to the temperature dependence of the equilibrium constant, the temperature and pressure dependency of the molar reaction Gibbs energy was determined to obtain an idea of the areas in the atmosphere where reaction (18) could occur spontaneously. The results are on display in Figure 6, and show a qualitative agreement with the earlier predictions by Natsheh et al., meaning that a decrease in pressure results in an increase in the Gibbs free energy and a decrease in temperature results in a more negative value for \( \Delta_rG_m \). The hydration process is thus most efficient at low temperatures.
Figure 5: The equilibrium constant as a function of temperature and height from sea level and normal pressures.

Because both the pressure and temperature decrease as one moves up in troposphere, these effects tend to cancel each other out. The reaction Gibbs energy stays roughly between $-8 \text{ kJ mol}^{-1}$ and $-12 \text{ kJ mol}^{-1}$, meaning that formation occurs throughout the troposphere. For example, in Figure 6, we see that the starting point at the right with $T = 280 \text{ K}$ and $p = 1 \text{ bar}$ is almost at the same level as the point in the tropopause with $p = 0.01 \text{ bar}$ and $T = 210 \text{ K}$. As a consequence of the linear dependency of the molar translational partition function on the pressure and equation (138), the pressure dependence of the Gibbs energy is predicted to be logarithmic.

In the stratosphere, the temperature begins to increase with increasing altitude due to ozone absorption, and so the hydrate formation will probably cease rather quickly. This effect is likely further enhanced by the radical drop in water concentration when one moves from the troposphere to the stratosphere as the water tends to freeze out. Thus, it is a safe assumption that in the stratosphere the first steps in the formation of the condensation nuclei do not follow reactions (18) and (19).

Finally, a study was also made on the temperature and pressure dependence of the reaction entropy. The results are shown in Figure 7 where one can see that the entropy is almost independent of the temperature, but does show variation with the pressure. Because $(\partial H/\partial p)_T = 0$ for an ideal gas, the whole pressure dependence of the Gibbs free energy arises from the entropy. As the temperature dependencies of the entropy and enthalpy are both small, the strong temperature dependence of the Gibbs free energy is caused by the linear $T$ dependence in its definition $G = H - TS$ - a fact that is strongly supported by the shape of the curves in Figure 6.
Figure 6: Molar Gibbs free energies for the reaction at different pressures and temperatures.

Figure 7: Molar entropies for the reaction at different pressures and temperatures.
3.6 Summary and Discussion

In this study, thermodynamic data have been calculated for the sulfuric acid monohydrate and the temperature and pressure dependencies of the thermodynamic properties have been investigated. The CCSD(T)-F12a method used to optimize the geometries of the reacting species and to calculate the ground state energies is by far more accurate than those used in previous works and has been shown to give results comparable with calculations done at the CCSD(T)/AVQZ level. Because of this, I conclude that the reaction energies are trustworthy, and recommend that future studies also make use of a similar higher-level calculation for the energies. The geometry optimizations confirmed the previously found ground state structures corresponding to chemical intuition, but I also found that there exists a low-lying stable configuration for the sulfuric acid monohydrate given by g2 in Figure 2, which should be accounted for in all future studies on this system. While this geometry has been reported by other studies, I feel that its significance has not been sufficiently underlined.

The harmonic frequencies used in this study were calculated with the DF-SCS-LMP2/AVTZ and CCSD-F12a/VDZ-F12 methods. Due to the neglect of anharmonicity, the ab initio wavenumbers usually show large discrepancies from experimental values with the high frequency OH stretches, but the DF-SCS-LMP2/AVTZ method showed a good agreement in the middle range of frequencies, outperforming even anharmonic calculations. This is most likely just the result of a fortuitous cancellation of errors. For the large amplitude motions, large discrepancies between both methods and the anharmonic calculations were again detected.

For the most important sources of anharmonicity, I made use of the anharmonic domain approximation, in which I isolated a few of the degrees of freedom and solved the complete vibrational problem variationally for these, while holding the rest of the molecule fixed to some geometry. The domains chosen were the two-dimensional space spanned by the wagging angle of the free hydrogen in the hydrogen bound water molecule and the HOSO torsional angle of the free OH group in the sulfuric acid, together with the three-dimensional space spanned by the internal coordinates of the bound water molecule. Results for the two-dimensional surface showed radical deviation from the harmonic case as the zero point energy almost doubled and in general the number of low-lying states decreased, causing large changes in the vibrational partition function. My results indicate that for a system with many energetically close minima, it is essential to employ an anharmonic treatment that takes all the minima into account. For large systems this might pose a problem, but according to my results, accurate values for the zero point energies and few of the lowest energy states can be obtained just from one-dimensional calculations. Introduction of additional couplings between the different degrees of freedom does not seem to add new low-lying states to the system, so a simple one-dimensional treatment is probably accurate enough for the calculation of thermodynamic properties in most cases.

The vibrational wavenumbers obtained from the three-dimensional calculation on complexated and free water molecules showed a good agreement with experimental results, though the lack of couplings to the other vibrational degrees of freedom was evident in the HOH bend and the OH stretch.
of the bonded hydrogen in the complexated water molecule. To account for the use of matrix isolation techniques in the experimental values, the effects of the Ar matrix were estimated by comparing the wavenumber shifts upon complexation for the different fundamental wavenumbers which were all found to be below 20 cm$^{-1}$. Because the vibrational wavefunction is less dependent on the high-wavenumber vibrations this accuracy is fine. To my best knowledge, approximations akin to the anharmonic domain approximation introduced in this study have not previously been used, at least in atmospheric computations. If this is true, then the good results reported in this study indicate that I have found a rather simple and efficient way to obtain large numbers of fundamental and overtone anharmonic states, with enough accuracy to be used in statistical thermodynamics calculations to obtain the thermodynamic properties. It should be noted, though, that to obtain more than just the few first vibrational states accurately, the domains must include all motions with strong couplings to each other.

Introduction of the anharmonic domains had a unifying effect on the thermodynamic data with the harmonic frequencies calculated either at the DF-SCS-LMP2/AVTZ method or the CCSD(T)-F12a/VDZ-F12a method. The values for reaction Gibbs energies, enthalpies, and zero point energies all were within 0.2 kJ mol$^{-1}$ of each other, and the difference in entropies was only 0.8 J mol$^{-1}$ K$^{-1}$. The uniformity in the enthalpies is mainly the result of fact that the CCSD(T)-F12a/VDZ-F12 method was used to calculate the reaction energies in each case while the entropies are more affected by the vibrational states and the anharmonic treatment. Compared with the rather chaotic results of the previous computational studies, the coherent results in this work indicate that I have discovered a more systematic way to estimate the thermodynamical data for this kind of relatively large atmospheric clusters.

The temperature dependence of the equilibrium constant was found to increase exponentially with decreasing temperature, resulting from the negative values for the reaction enthalpies and entropies. From the statistical mechanical formulae derived for this work it was also found that the enthalpy could be approximated as a constant all the way up to 230 K where the error was about 3 %. At 210 K the error was found to be 6 %. The pressure and temperature dependences of the reaction Gibbs free energy and entropy were also mapped and from them, I predicted that the reaction Gibbs energy would have a roughly constant value throughout the troposphere, resulting from the cancellation of the different factors caused by the decreasing temperature and pressure. The entropy was found to be almost independent of temperature, but it did show a positive pressure dependence that was reflected in the negative pressure dependence of the Gibbs free energy.

At present, my calculations ignore completely the second sulfuric acid geometry and it would be interesting to take it into account propably by an introduction of some kind of anharmonic domain. At the very least, this would have to include the two torsional motions of the OH groups. Because the harmonic calculations show poor agreement with the high frequency OH stretches, these should propably be calculated with the anharmonic domains as well. According to the normal mode analysis, the free OH group of the sulfuric acid in the monohydrate functions as a rather isolated system, so a good choice for a two-dimensional domain might be the OH bond length.
and the SOH angle. The analysis done on the anharmonic couplings of the sulfuric acid monohydrate by Miller et al.\textsuperscript{30} also indicates that the coupling is strongest between these two modes. Their study also shows that the bound OH is strongly coupled to many different motions, but perhaps a one-dimensional domain would be enough for the accurate prediction of the zero point energy. Because the vibrational states are thousands of wavenumbers apart, the overtones have negligible effect on the partition function.

As shown by the results of Miller et al., large anharmonic effects are also present in the remaining harmonically treated large amplitude motions.\textsuperscript{30} This problem would be solved by either using some kind of anharmonic corrections to the calculated DF-SCS-LMP2/AVTZ values or by methods such as the CC-VSCF. At least with the latter, the problem is that it is difficult to calculate the vibrational overtones,\textsuperscript{30} which are essential for the large amplitude motions.

The systematic approach represented in this work can clearly be extended to other similar complexes. Possible avenues for future research would be to use this method for the sulfuric acid ammonia complex, and perhaps even complexes containing three molecules such as the sulfuric acid dihydrate or \((\text{H}_2\text{SO}_4)(\text{H}_2\text{O})(\text{NH}_3)\). In the case of the larger complexes, the geometry optimization might be too computationally expensive, but as can be seen in Appendix II, the geometry changes only a little when moving from CCSD-F12a to CCSD(T)-F12a and so it would be possible to do the optimization at the CCSD-F12a level and use those geometries for the CCSD(T)-F12a electronic energy calculation. This approach of calculating the geometry at a lower level of theory has been successfully implemented in, for example, many composite extrapolation procedures such as the G2 and G3 methods.\textsuperscript{65} If this extension proves possible, then we might be able to estimate simple scaling factors separately for the large amplitude motions, middle range wavenumbers and high-wavenumber OH stretches, which would make it possible to accurately treat even larger systems where the anharmonic domain approximation itself becomes too unwieldy.
References


71


72


Appendix I: Equilibrium geometries of the different sulfuric acid conformers and water. The angles are given in degrees and the bond lengths in ångströms.

<table>
<thead>
<tr>
<th></th>
<th>H₂SO₄</th>
<th>H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DF-SCS-LMP2/AVDZ g₁</td>
<td>DF-SCS-LMP2/AVTZ g₁</td>
</tr>
<tr>
<td>r1 / Å</td>
<td>0.976 0.975</td>
<td>0.970 0.969</td>
</tr>
<tr>
<td>r2 / Å</td>
<td>1.650 1.650</td>
<td>1.603 1.602</td>
</tr>
<tr>
<td>r3 / Å</td>
<td>1.466 1.467</td>
<td>1.433 1.434</td>
</tr>
<tr>
<td>a1</td>
<td>106.6 106.0</td>
<td>107.9 107.4</td>
</tr>
<tr>
<td>a2</td>
<td>108.7 107.0</td>
<td>108.8 107.5</td>
</tr>
<tr>
<td>a3</td>
<td>125.3 125.1</td>
<td>124.3 124.1</td>
</tr>
<tr>
<td>a4</td>
<td>101.3 101.3</td>
<td>101.9 101.7</td>
</tr>
<tr>
<td>d₉O₆SO₂H₁</td>
<td>83.5 97.4</td>
<td>83.7 94.2</td>
</tr>
<tr>
<td>d₉H₇O₆SO₂</td>
<td>83.5 -97.6</td>
<td>83.7 -102.3</td>
</tr>
<tr>
<td>r / Å</td>
<td>0.962</td>
<td>0.956</td>
</tr>
<tr>
<td>a</td>
<td>104.2</td>
<td>104.5</td>
</tr>
</tbody>
</table>

a Kuczkowski et al. 36
b at the CCSD(T)(fc)/V(T+d)Z level by Demaison et al. 147
c at the MP2/6-311++G(2d,2p) level by Beichert and Schrems 39
d at the PW91/TZP level by Al Natsheh et al. 9
e at the B3LYP/6-311++G(2d,2p) level by Bandy and Ianni 10
f Benedict et al. 38

H₂O

<table>
<thead>
<tr>
<th></th>
<th>DF-SCS-LMP2/AVDZ</th>
<th>DF-SCS-LMP2/AVTZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>r / Å</td>
<td>0.962</td>
<td>0.956</td>
</tr>
<tr>
<td>a</td>
<td>104.2</td>
<td>104.5</td>
</tr>
</tbody>
</table>

H₂O

<table>
<thead>
<tr>
<th></th>
<th>DF-SCS-LMP2/AVDZ</th>
<th>DF-SCS-LMP2/AVTZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>r / Å</td>
<td>0.962</td>
<td>0.956</td>
</tr>
<tr>
<td>a</td>
<td>104.2</td>
<td>104.5</td>
</tr>
</tbody>
</table>

f Hoy and Bunker 144
b Cook et al. 145
i at the MP2/AVTZ level by Kim et al. 37
j at the CCSD(T)/DZP level by Kim et al. 37
k at the CCSD(T)/AVQZ level by Lane and Kjaergaard 120
Appendix II: Equilibrium geometries of the different sulfuric acid monohydrate conformers. Angles are given in degrees and bond lengths are given in ångströms.

| r1 / Å | r2 / Å | r3 / Å | r4 / Å | r5 / Å | r6 / Å | r7 / Å | r8 / Å | r9 / Å | r10 / Å | a1   | a2   | a3   | a4   | a5   | a6   | a7   | a8   | d_{O6S3O2H1} | d_{H6O6S3O2} | d_{O6H6O6S3} | d_{H6H6H6S3} |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|----------------|----------------|----------------|----------------|
| 0.975  | 1.652  | 1.466  | 1.473  | 1.625  | 0.995  | 1.779  | 0.971  | 0.966  | 2.324  | 106.7 | 108.6 | 123.8 | 109.5 | 102.1 | 107.0 | 165.5 | 105.4         | 81.6           | 82.6           | 10.8           |
| 0.976  | 1.658  | 1.466  | 1.471  | 1.626  | 0.995  | 1.758  | 0.970  | 0.967  | 2.683  | 106.4 | 108.9 | 124.0 | 110.0 | 101.6 | 106.8 | 171.1 | 104.7         | 72.2           | 77.1           | 8.8            |
| 0.975  | 1.652  | 1.459  | 1.483  | 1.626  | 0.992  | 1.715  | 0.970  | 0.966  | 2.322  | 106.6 | 105.8 | 123.8 | 108.7 | 102.3 | 107.2 | 164.7 | 105.8         | -89.9          | 89.3           | 4.8            |
| 0.976  | 1.645  | 1.467  | 2.553  | 2.557  | 0.992  | 1.717  | 0.970  | 0.966  | 2.663  | 106.0 | 105.7 | 122.7 | 108.6 | 103.1 | 108.6 | 166.3 | 106.3         | -87.9          | 96.3           | 6.4            |
| 0.975  | 1.598  | 1.426  | 1.449  | 1.579  | 0.990  | 1.717  | 0.967  | 0.965  | 2.665  | 108.0 | 105.7 | 122.7 | 123.5 | 103.7 | 108.6 | 166.3 | 106.3         | 82.5           | 85.2           | 11.6           |
| 0.976  | 1.580  | 1.416  | 1.449  | 1.579  | 0.990  | 1.717  | 0.966  | 0.965  | 2.665  | 108.0 | 105.7 | 122.7 | 123.5 | 103.7 | 108.6 | 166.3 | 106.3         | 72.7           | 78.9           | 28.8           |
| 1.598  | 1.416  | 1.409  | 2.519  | 2.528  | 0.990  | 1.717  | 0.966  | 0.965  | 2.665  | 108.0 | 105.7 | 122.7 | 123.5 | 103.7 | 108.6 | 166.3 | 106.3         | -87.9          | 96.3           | 6.4            |
| 1.580  | 1.416  | 1.417  | 2.519  | 2.528  | 0.990  | 1.717  | 0.966  | 0.965  | 2.665  | 108.0 | 105.7 | 122.7 | 123.5 | 103.7 | 108.6 | 166.3 | 106.3         | 82.5           | 85.2           | 11.6           |
| 1.414  | 1.414  | 1.424  | 2.396  | 2.386  | 0.994  | 1.677  | 0.966  | 0.959  | 2.140  | 108.4 | 108.2 | 102.2 | 123.1 | 103.2 | 108.9 | 164.4 | 106.0         | 84.7           | 89.4           | 12.6           |
| 1.414  | 1.414  | 1.424  | 2.396  | 2.386  | 0.994  | 1.677  | 0.966  | 0.959  | 2.140  | 108.4 | 108.2 | 102.2 | 123.1 | 103.2 | 108.9 | 164.4 | 106.0         | 84.7           | 89.4           | 12.6           |
| 1.410  | 1.410  | 1.424  | 2.396  | 2.386  | 0.994  | 1.677  | 0.966  | 0.959  | 2.140  | 108.4 | 108.2 | 102.2 | 123.1 | 103.2 | 108.9 | 164.4 | 106.0         | 106.2          | 106.0          | 12.6           |
| 1.410  | 1.410  | 1.424  | 2.396  | 2.386  | 0.994  | 1.677  | 0.966  | 0.959  | 2.140  | 108.4 | 108.2 | 102.2 | 123.1 | 103.2 | 108.9 | 164.4 | 106.0         | 106.2          | 106.0          | 12.6           |
| 1.410  | 1.410  | 1.424  | 2.396  | 2.386  | 0.994  | 1.677  | 0.966  | 0.959  | 2.140  | 108.4 | 108.2 | 102.2 | 123.1 | 103.2 | 108.9 | 164.4 | 106.0         | 106.2          | 106.0          | 12.6           |
Appendix III: Harmonic wavenumbers for sulfuric acid and water in cm\(^{-1}\). The VDZ-F12 basis has been denoted by VDZ.

\[
\begin{array}{|l|c|c|c|c|c|c|c|c|c|}
\hline
& \text{DF-SCS-LMP2/AVTZ} & \text{CCSD-F12A/VDZ} & \text{CCSD(T)-F12A/VDZ} & \text{experimental} & \text{other calculations} \\
\hline
\text{H}_2\text{SO}_4 \text{OSOH symmetric torsion} & 247 & 45 & 246 & 237 & 238 & 229 & 222 & 380 & 221 & 105 \\
\text{OSOH asymmetric torsion} & 329 & 323 & 321 & 311 & 314 & 304 & 303 & 379 & 321 & 323 \\
\text{(OH)S(OH) bend} & 370 & 366 & 391 & 381 & 344 & 331 & 338 & 459 & 341 & 349 \\
\text{O-S=O bend} & 433 & 415 & 453 & 443 & 408 & 395 & 401 & 489 & 413 & 398 \\
\text{S=O} \text{rock} & 490 & 493 & 518 & 506 & 459 & 443 & 450 & 525 & 463 & 466 \\
\text{OSOH asymmetric torsion} & 535 & 534 & 569 & 553 & 551 & 550 & 550 & 497 & 499 & 499 \\
\text{S=O wag} & 545 & 545 & 581 & 565 & 509 & 487 & 500 & 480 & 513 & 514 \\
\text{S-OH symmetric stretch} & 817 & 815 & 883 & 853 & 834 & 834 & 831 & 722 & 696 & 736 \\
\text{S-OH asymmetric stretch} & 867 & 869 & 933 & 906 & 891 & 883 & 882 & 776 & 750 & 793 \\
\text{symmetric SOH bend} & 1175 & 1163 & 1180 & 1167 & 1138 & 1141 & 1115 & 1109 & 1115 & 1118 \\
\text{asymmetric SOH bend} & 1188 & 1188 & 1204 & 1190 & 1157 & 1159 & 1127 & 1122 & 1126 & 1128 \\
\text{S=O} \text{symmetric stretch} & 1218 & 1216 & 1275 & 1247 & 1220 & 1223 & 1224 & 1145 & 1125 & 1142 \\
\text{S=O} \text{asymmetric stretch} & 1467 & 1465 & 1525 & 1499 & 1465 & 1450 & 1456 & 1392 & 1356 & 1388 \\
\text{asymmetric OH stretch} & 3768 & 3779 & 3832 & 3785 & 3669 & 3610 & 3610 & 3626 & 3600 & 3642 \\
\text{symmetric OH stretch} & 3772 & 3784 & 3836 & 3789 & 3669 & 3610 & 3610 & 3626 & 3603 & 3646 \\
\hline
\text{H}_2\text{O} \text{HOH bend} & 1646 & 1663 & 1652 & 1648 & 1677 & 1648 & 1624 \\
\text{symmetric OH stretch} & 3811 & 3870 & 3834 & 3832 & 3894 & 3832 & 3848 \\
\text{asymmetric OH stretch} & 3933 & 3976 & 3942 & 3943 & 4029 & 3942 & 3969 \\
\hline
\end{array}
\]

\(^a\) Hintze et. al. \(^{158}\)
\(^b\) Chackalackal and Stafford \(^{40}\)
\(^c\) Stopperka and Kilz \(^{150}\)
\(^d\) at the BLYP/6-311++G(3df,3pd) level by Nadykto et al. \(^{29}\)
\(^e\) at the BLYP/AVTZ level by Nadykto et al. \(^{29}\)
\(^f\) at the PW91/AVTZ level by Nadykto et al. \(^{29}\)
\(^g\) anharmonic wavenumbers at the CC-VSCF/MPI2-TZP level by Miller et al. \(^{30}\)
\(^h\) at the PW91/TZP level by Al Natsheh et al. \(^9\)
\(^i\) Experimental harmonic vibrational wavenumbers by Benedict et al. \(^{38}\)
\(^j\) at the CCSD(T)/DZP level by Kim et al. \(^{37}\)
\(^k\) at the CCSD(T)/(13,8,4,2/8,4,2) level by Kim et al. \(^{37}\)
\(^l\) at the MP2/AVTZ level by Kim et al. \(^{37}\)
### Appendix IV: Harmonic wavenumbers for sulfuric acid monohydrate in cm\(^{-1}\)

<table>
<thead>
<tr>
<th></th>
<th>DF-SCS-LMP2/AVTZ</th>
<th>CCSD-F12A/VDZ</th>
<th>experimental</th>
<th>other calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>intmolec. mode</td>
<td>g1, g2, g3</td>
<td>g1</td>
<td>g1 (e), g1 (b)</td>
<td>g1 (e), g1 (d), g1 (e), g1 (f), g1 (g), g2 (e), g3 (e)</td>
</tr>
<tr>
<td>intmolec. mode</td>
<td>47, 19, 28</td>
<td>119, 119, 146</td>
<td>49, 49, 49, 104</td>
<td>48, 55, 167, 155, 160</td>
</tr>
<tr>
<td>intmolec. mode</td>
<td>120, 53, 123</td>
<td>128</td>
<td>122, 209, 218, 529</td>
<td>121, 228, 228, 204</td>
</tr>
<tr>
<td>intmolec. mode</td>
<td>189, 117, 191</td>
<td>178</td>
<td>228, 229, 255, 331</td>
<td>258, 263, 249</td>
</tr>
<tr>
<td>intmolec. mode</td>
<td>211, 214, 220</td>
<td>212</td>
<td>246</td>
<td></td>
</tr>
<tr>
<td>free HOSO torsion</td>
<td>264, 271, 229</td>
<td>318</td>
<td>253, 244, 260, 397</td>
<td>284, 286, 304</td>
</tr>
<tr>
<td>bound HOSO torsion</td>
<td>312, 311, 315</td>
<td>318</td>
<td>338, 327, 337, 529</td>
<td>339, 345, 359</td>
</tr>
<tr>
<td>(OH)S(OH) bend</td>
<td>319, 371, 371</td>
<td>395</td>
<td>303, 394, 311, 402</td>
<td>411, 412, 407</td>
</tr>
<tr>
<td>O-S=O bend</td>
<td>410, 398, 417</td>
<td>442</td>
<td>403, 389, 402, 486</td>
<td>411, 412, 407</td>
</tr>
<tr>
<td>intmolec. mode</td>
<td>462, 432, 454</td>
<td>484</td>
<td>476, 462, 480, 540</td>
<td>491, 488, 489</td>
</tr>
<tr>
<td>S=O (2) rock</td>
<td>519, 511, 524</td>
<td>547</td>
<td>504, 486, 500, 536</td>
<td>509, 509, 510</td>
</tr>
<tr>
<td>OSO bend</td>
<td>544, 540, 540</td>
<td>577</td>
<td>517, 497, 515, 514</td>
<td>527, 525, 530</td>
</tr>
<tr>
<td>S=O (2) wag</td>
<td>551, 548, 553</td>
<td>588</td>
<td>537, 528, 584, 476</td>
<td>635, 624, 646</td>
</tr>
<tr>
<td>intmolec. mode</td>
<td>762, 814, 741</td>
<td>797</td>
<td>734, 709, 749, 787</td>
<td>761, 759, 768</td>
</tr>
<tr>
<td>S-OH symmetric stretch</td>
<td>828, 858, 828</td>
<td>893, 839, 834</td>
<td>816, 792, 843, 834</td>
<td>856, 853, 848</td>
</tr>
<tr>
<td>S-OH asymmetric stretch</td>
<td>904, 899, 904</td>
<td>969, 929, 887</td>
<td>828, 808, 882, 868</td>
<td>917, 920, 930</td>
</tr>
<tr>
<td>free SOH bend</td>
<td>1179, 1183, 1172</td>
<td>1190</td>
<td>1179, 1109, 1112, 1153</td>
<td>1131, 1131, 1121</td>
</tr>
<tr>
<td>S=O (2) symmetric stretch</td>
<td>1209, 1217, 1200</td>
<td>1261, 1205</td>
<td>1134, 1124, 1128, 1172</td>
<td>1140, 1144, 1144</td>
</tr>
<tr>
<td>bound SOH bend</td>
<td>1355, 1330, 1348</td>
<td>1382</td>
<td>1307, 1281, 1313, 1336</td>
<td>1343, 1343, 1351</td>
</tr>
<tr>
<td>S=O (2) asymmetric stretch</td>
<td>1467, 1461, 1468</td>
<td>1512, 1500</td>
<td>1412, 1401, 1437, 1436, 1467, 1469, 1452</td>
<td></td>
</tr>
<tr>
<td>HOH bend on H(2)O</td>
<td>1642, 1656, 1634</td>
<td>1658, 1600</td>
<td>1595, 1595, 1593, 1573, 1607, 1606, 1601</td>
<td></td>
</tr>
<tr>
<td>bound OH stretch on H(2)SO(_4)</td>
<td>3320, 3322, 3361</td>
<td>3332, 2722</td>
<td>2989, 2970, 2817, 2973, 2771, 2778, 2757</td>
<td></td>
</tr>
<tr>
<td>H(2)O symmetric stretch</td>
<td>3769, 3770, 3773</td>
<td>3810</td>
<td>3600, 3580, 3562, 3593, 3508, 3509, 3495</td>
<td></td>
</tr>
<tr>
<td>free OH stretch on H(2)SO(_4)</td>
<td>3775, 3788, 3781</td>
<td>3839</td>
<td>3632, 3609, 3650, 3628, 3638, 3631, 3638</td>
<td></td>
</tr>
<tr>
<td>H(2)O asymmetric stretch</td>
<td>3908, 3905, 3910</td>
<td>3944</td>
<td>3697, 3745, 3736, 3720, 3757, 3738, 3730, 3734, 3731</td>
<td></td>
</tr>
</tbody>
</table>

\(a\) Rozenberg and Loewenschuss\(^{44}\)

\(b\) Givan et. al.\(^{41}\)

\(c\) at the BLYP/6-311++G(3df,3pd) level by Nadykto et al.\(^{29}\)

\(d\) at the BLYP/AVTZ level by Nadykto et al.\(^{29}\)

\(e\) anharmonic wavenumbers at the CC-VSCF/MP2-TZP level by Miller et al.\(^{30}\)

\(f\) at the PW91/TZP level by Al Natsheh et al.\(^{9}\)

\(g\) at the PW91/AVTZ level by Nadykto et al.\(^{29}\)
Appendix V: Vibrational states of water in sulfuric acid monohydrate. All calculations have been done at the CCSD(T)-F12a/VDZ-F12 level

<table>
<thead>
<tr>
<th>state</th>
<th>wavenumber / cm⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0],[0],[1]</td>
<td>1589.2</td>
</tr>
<tr>
<td>[0],[0],[2]</td>
<td>3146.1</td>
</tr>
<tr>
<td>[0],[1],[0],[0] (85 %) +</td>
<td>[1],[0],[0],[0] (13 %)</td>
</tr>
<tr>
<td>[1],[0],[0],[0] (85 %) +</td>
<td>[0],[1],[0],[0] (13 %)</td>
</tr>
<tr>
<td>[0],[0],[3]</td>
<td>4669.5</td>
</tr>
<tr>
<td>[0],[1],[1],[0] (84 %) +</td>
<td>[1],[0],[0],[1] (11 %)</td>
</tr>
<tr>
<td>[1],[0],[0],[1] (85 %) +</td>
<td>[0],[1],[0],[1] (12 %)</td>
</tr>
<tr>
<td>[0],[0],[4] (81 %) +</td>
<td>[0],[0],[5] (11 %)</td>
</tr>
<tr>
<td>[0],[1],[2] (81 %) +</td>
<td>[1],[0],[2] (9 %)</td>
</tr>
<tr>
<td>[1],[0],[0],[2] (82 %) +</td>
<td>[0],[1],[0],[2] (11 %)</td>
</tr>
<tr>
<td>[0],[2],[0]</td>
<td>7019.2</td>
</tr>
<tr>
<td>[2],[0],[0],[0] (67 %) +</td>
<td>[1],[1],[0],[0] (24 %)</td>
</tr>
<tr>
<td>[1],[1],[0],[0] (67 %) +</td>
<td>[2],[0],[0],[0] (28 %)</td>
</tr>
<tr>
<td>[0],[0],[5] (69 %) +</td>
<td>[0],[0],[6] (20 %)</td>
</tr>
<tr>
<td>[0],[2],[1]</td>
<td>8548.8</td>
</tr>
<tr>
<td>[2],[0],[0],[1] (61 %) +</td>
<td>[1],[1],[0],[1] (26 %)</td>
</tr>
<tr>
<td>[1],[1],[0],[1] (62 %) +</td>
<td>[2],[0],[0],[1] (30 %)</td>
</tr>
<tr>
<td>[0],[0],[6] (52 %) +</td>
<td>[0],[0],[7] (32 %) +</td>
</tr>
<tr>
<td>[0],[3],[0]</td>
<td>10302.3</td>
</tr>
<tr>
<td>[0],[0],[8] (39 %) +</td>
<td>[0],[0],[7] (38 %) +</td>
</tr>
<tr>
<td>[3],[0],[0] (70 %) +</td>
<td>[2],[1],[0],[0] (14 %)</td>
</tr>
<tr>
<td>[1],[2],[0] (65 %) +</td>
<td>[3],[0],[0] (15 %) +</td>
</tr>
<tr>
<td>[2],[1],[0] (70 %) +</td>
<td>[1],[2],[0] (19 %)</td>
</tr>
<tr>
<td>[0],[5],[1]</td>
<td>11802.4</td>
</tr>
<tr>
<td>[3],[0],[1] (59 %) +</td>
<td>[2],[1],[1],[1] (14 %)</td>
</tr>
<tr>
<td>[1],[2],[1] (60 %) +</td>
<td>[3],[0],[1],[1] (18 %)</td>
</tr>
<tr>
<td>[2],[1],[1] (67 %) +</td>
<td>[1],[2],[1] (18 %)</td>
</tr>
<tr>
<td>[0],[4],[0] (85 %) +</td>
<td>[0],[5],[0] (10 %)</td>
</tr>
<tr>
<td>[4],[0],[0] (77 %)</td>
<td>13850.9</td>
</tr>
<tr>
<td>[1],[3],[0] (71 %) +</td>
<td>[2],[2],[0] (11 %)</td>
</tr>
<tr>
<td>[3],[1],[0] (50 %) +</td>
<td>[2],[2],[0] (23 %) +</td>
</tr>
<tr>
<td>[2],[2],[0] (56 %) +</td>
<td>[3],[1],[0] (32 %)</td>
</tr>
<tr>
<td>[0],[4],[1] (71 %) +</td>
<td>[0],[5],[1] (13 %)</td>
</tr>
<tr>
<td>[1],[3],[1] (65 %) +</td>
<td>[2],[2],[1] (9 %)</td>
</tr>
<tr>
<td>[3],[1],[1] (42 %) +</td>
<td>[2],[2],[1] (23 %) +</td>
</tr>
<tr>
<td>[2],[2],[1] (51 %) +</td>
<td>[3],[1],[1] (32 %)</td>
</tr>
</tbody>
</table>
Appendix VI: Experimental and calculated CCSD(T)-F12a/VDZ-F12 level vibrational states of the water monomer and the energy differences ΔE between computational and experimental values for my results and those of others. All values are given in cm⁻¹.

<table>
<thead>
<tr>
<th></th>
<th>calc.</th>
<th>exp.ᵃ</th>
<th>ΔEᵇ</th>
<th>ΔEᶜ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[00]⁺[1]</td>
<td>1594.5</td>
<td>1594.8</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>[00]⁺[2]</td>
<td>3153.8</td>
<td>3151.6</td>
<td>2.2</td>
</tr>
<tr>
<td>[10]⁻[0]</td>
<td>3651.9</td>
<td>3657.1</td>
<td>-5.2</td>
<td>-5.6</td>
</tr>
<tr>
<td>[10]⁻[0]</td>
<td>3761.0</td>
<td>3755.9</td>
<td>5.1</td>
<td>0.5</td>
</tr>
<tr>
<td>[10]⁺[1]</td>
<td>5224.1</td>
<td>5235.0</td>
<td>-10.9</td>
<td>-8.2</td>
</tr>
<tr>
<td>[10]⁻[1]</td>
<td>5335.1</td>
<td>5331.3</td>
<td>3.8</td>
<td>1.9</td>
</tr>
<tr>
<td>[10]⁺[2]</td>
<td>6760.3</td>
<td>6775.1</td>
<td>-14.8</td>
<td>-8.2</td>
</tr>
<tr>
<td>[10]⁻[2]</td>
<td>6874.9</td>
<td>6871.5</td>
<td>3.4</td>
<td>4.9</td>
</tr>
<tr>
<td>[20]⁻[0]</td>
<td>7195.4</td>
<td>7201.5</td>
<td>-6.1</td>
<td>-8.9</td>
</tr>
<tr>
<td>[20]⁻[0]</td>
<td>7250.8</td>
<td>7249.8</td>
<td>1.0</td>
<td>-5.4</td>
</tr>
<tr>
<td>[11]⁻[0]</td>
<td>7454.9</td>
<td>7445.1</td>
<td>9.8</td>
<td>0.9</td>
</tr>
<tr>
<td>[20]⁺[1]</td>
<td>8746.0</td>
<td>8761.6</td>
<td>-15.6</td>
<td>-13.2</td>
</tr>
<tr>
<td>[20]⁻[1]</td>
<td>8803.1</td>
<td>8807.0</td>
<td>-3.9</td>
<td>-5.6</td>
</tr>
<tr>
<td>[11]⁻[1]</td>
<td>9008.2</td>
<td>9000.1</td>
<td>8.1</td>
<td>3.4</td>
</tr>
<tr>
<td>[30]⁻[0]</td>
<td>10597.4</td>
<td>10599.7</td>
<td>-2.3</td>
<td>-10.4</td>
</tr>
<tr>
<td>[30]⁻[0]</td>
<td>10613.9</td>
<td>10613.4</td>
<td>0.5</td>
<td>-9.6</td>
</tr>
<tr>
<td>[21]⁻[0]</td>
<td>10870.9</td>
<td>10868.9</td>
<td>2.0</td>
<td>-8.0</td>
</tr>
<tr>
<td>[21]⁻[0]</td>
<td>11050.0</td>
<td>11032.4</td>
<td>17.6</td>
<td>3.2</td>
</tr>
<tr>
<td>[40]⁻[0]</td>
<td>13830.3</td>
<td>13828.3</td>
<td>2.0</td>
<td>-12.9</td>
</tr>
<tr>
<td>[40]⁻[0]</td>
<td>13833.8</td>
<td>13830.9</td>
<td>2.9</td>
<td>-12.4</td>
</tr>
<tr>
<td>[31]⁻[0]</td>
<td>14221.2</td>
<td>14221.1</td>
<td>0.1</td>
<td>-13.0</td>
</tr>
<tr>
<td>[31]⁻[0]</td>
<td>14331.4</td>
<td>14318.8</td>
<td>12.6</td>
<td>-6.0</td>
</tr>
<tr>
<td>[22]⁻[0]</td>
<td>14562.5</td>
<td>14536.9</td>
<td>25.6</td>
<td>5.0</td>
</tr>
<tr>
<td>[50]⁻[0]</td>
<td>16903.3</td>
<td>16898.4</td>
<td>4.9</td>
<td>-12.7</td>
</tr>
<tr>
<td>[50]⁻[0]</td>
<td>16903.9</td>
<td>16898.8</td>
<td>5.1</td>
<td>-12.4</td>
</tr>
</tbody>
</table>

ᵃ experimental value by Bykov et al.⁴²
ᵇ at the CCSD(T)/AVQZ level by Salmi et al.¹³¹
ᶜ at the CCSD(T)/AVTZ level by Salmi et al.¹³¹
Appendix VII: Vibrational states for the one- and two-dimensional treatment of the large amplitude motions.

All values are given in cm$^{-1}$

<table>
<thead>
<tr>
<th>2-dim.</th>
<th>1-dim.</th>
<th>CCSD(T)-F12/VDZ-F12</th>
<th>DF-SCS-LMP2/AVTZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>wagging</td>
<td>torsion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>torsion</td>
</tr>
<tr>
<td>297</td>
<td>301</td>
<td>156</td>
<td>144</td>
</tr>
<tr>
<td>332</td>
<td>333</td>
<td>188</td>
<td>403</td>
</tr>
<tr>
<td>535</td>
<td>539</td>
<td>395</td>
<td>545</td>
</tr>
<tr>
<td>557</td>
<td>560</td>
<td>485</td>
<td>638</td>
</tr>
<tr>
<td>598</td>
<td>592</td>
<td>656</td>
<td>778</td>
</tr>
<tr>
<td>627</td>
<td>629</td>
<td>839</td>
<td>844</td>
</tr>
<tr>
<td>707</td>
<td>702</td>
<td>1046</td>
<td>962</td>
</tr>
<tr>
<td>792</td>
<td>734</td>
<td>1267</td>
<td>1022</td>
</tr>
<tr>
<td>796</td>
<td>794</td>
<td>1114</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>798</td>
<td>1214</td>
<td></td>
</tr>
<tr>
<td>841</td>
<td>801</td>
<td>1278</td>
<td></td>
</tr>
<tr>
<td>890</td>
<td>826</td>
<td>1432</td>
<td></td>
</tr>
<tr>
<td>939</td>
<td>888</td>
<td>1451</td>
<td></td>
</tr>
<tr>
<td>959</td>
<td>934</td>
<td></td>
<td></td>
</tr>
<tr>
<td>982</td>
<td>940</td>
<td></td>
<td></td>
</tr>
<tr>
<td>999</td>
<td>966</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1006</td>
<td>983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1035</td>
<td>1001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1060</td>
<td>1030</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1062</td>
<td>1032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1122</td>
<td>1033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1130</td>
<td>1060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1168</td>
<td>1118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1178</td>
<td>1122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1184</td>
<td>1150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1189</td>
<td>1172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1229</td>
<td>1178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1245</td>
<td>1190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1245</td>
<td>1202</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1248</td>
<td>1210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1272</td>
<td>1239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1301</td>
<td>1242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1331</td>
<td>1262</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1340</td>
<td>1270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1346</td>
<td>1294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1372</td>
<td>1302</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1391</td>
<td>1329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1403</td>
<td>1357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1410</td>
<td>1371</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1429</td>
<td>1384</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1436</td>
<td>1403</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1442</td>
<td>1411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1452</td>
<td>1416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1484</td>
<td>1434</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>